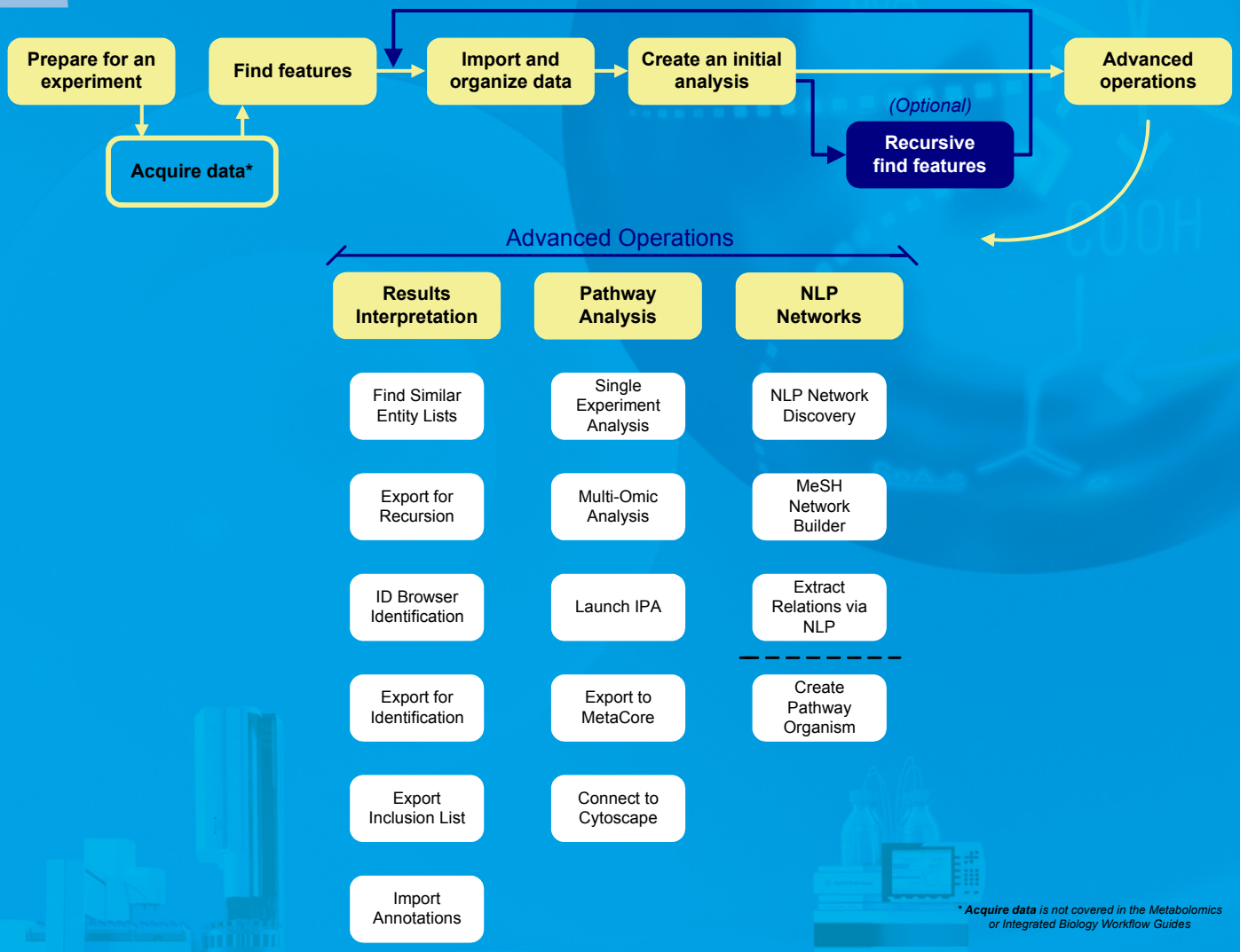
Integrated Biology with Agilent Mass Profiler Professional

Workflow Guide



* ***Acquire data*** *is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Agilent Technologies

Notices

© Agilent Technologies, Inc. 2013

No part of this manual may be reproduced in any form or by any means (including electronic storage and retrieval or translation into a foreign language) without prior agreement and written consent from Agilent Technologies, Inc. as governed by United States and international copyright laws.

Manual Part Number

5991-1909EN

Edition

Revision A, June 2013

Printed in USA

Agilent Technologies, Inc.
5301 Stevens Creek Blvd.
Santa Clara, CA 95051

Acknowledgements

Microsoft is either a registered trademark or trademark of Microsoft Corporation in the United States and/or other countries.

Adobe is a trademark of Adobe Systems Incorporated.

Warranty

The material contained in this document is provided "as is," and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights

If software is for use in the performance of a U.S. Government prime contract or subcontract, Software is delivered and licensed as "Commercial computer software" as defined in DFAR 252.227-7014 (June 1995), or as a "commercial item" as defined in FAR 2.101(a) or as "Restricted computer software" as defined in FAR 52.227-19 (June 1987) or any equivalent agency regulation or contract clause. Use, duplication or disclosure of Software is subject to Agilent Technologies' standard commercial license terms, and non-DOD Departments and Agencies of the U.S. Government will receive no greater than Restricted Rights as defined in FAR 52.227-19(c)(1-2) (June 1987). U.S. Government users will receive no greater than Limited Rights as defined in FAR 52.227-14 (June 1987) or DFAR 252.227-7015 (b)(2) (November 1995), as applicable in any technical data.

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

Contents

1 Before you begin 5
Introduction 6
Required items 8
Compliance 10

2 Working with Mass Profiler Professional 11
Where is MPP used in your experiment? 12
What is the metabolomics workflow? 13
Advanced operations covered in the MPP workflow guides 16
Using Mass Profiler Professional 17

3 Example experiments 19
Features of the example mass spectrometry experiments 20
Features of the example array experiment 22
Creating an expression analysis using the sample array experiment 23

4 Integrated Biology operations 35
Overview of operations 36
Results Interpretation 38
Pathway Analysis 61
NLP Networks 92

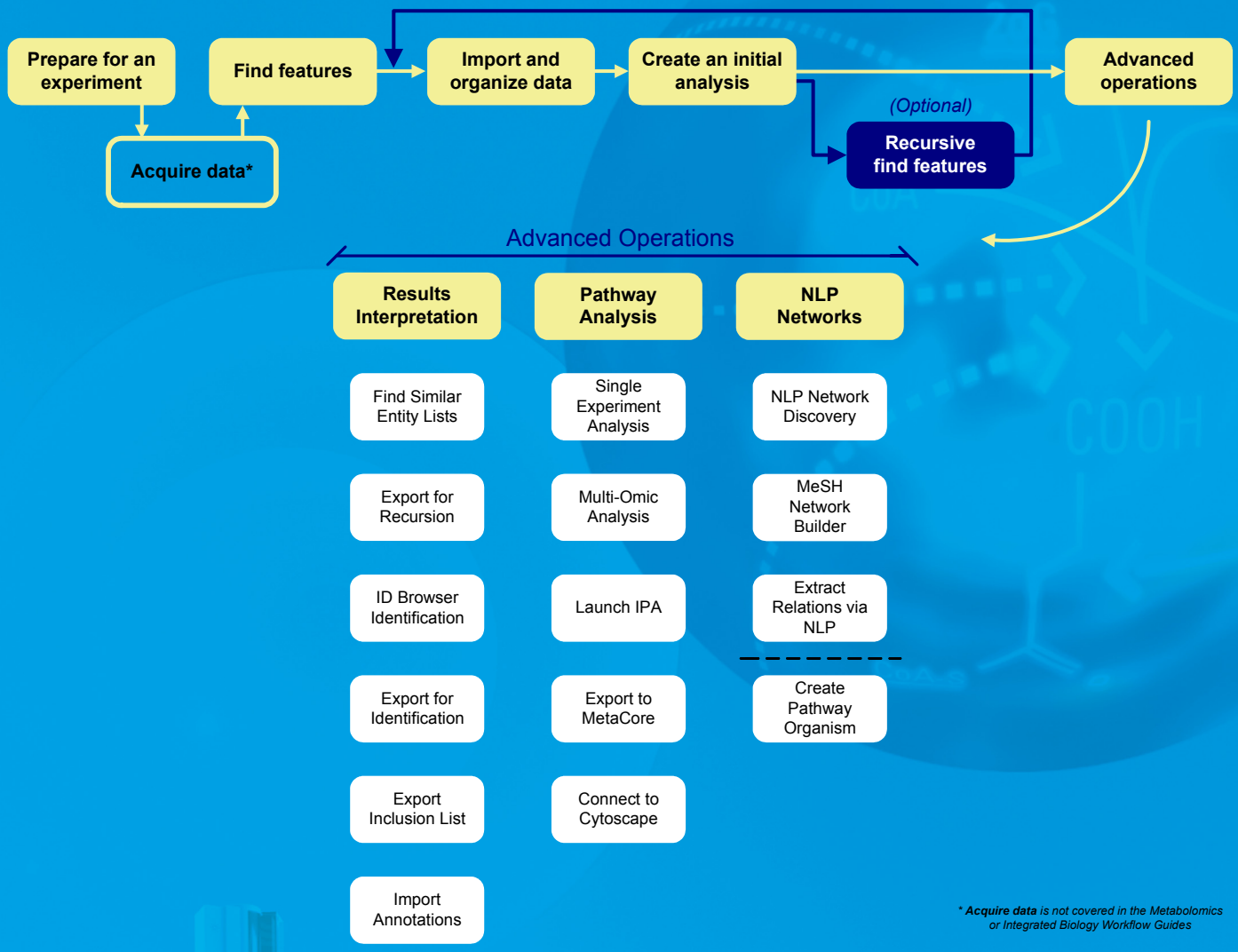
5 Reference information 109
Definitions 110
References 120

What's new in Revision A

- This workflow guide is a complementary guide to the *Agilent Metabolomics Workflow - Discovery Workflow Guide* (Agilent publication 5990-7067EN, Revision B) The Metabolomics Workflow presents steps that precede the operations used in the Integrated Biology Workflow.
- The Mass Profiler Professional wizard and workflow images are based on version 12.05.
- Formatting of text that appears in the left-hand margin helps guide you through the operations.
- Operations are illustrated with flow charts that show you how the wizards are navigated based on your experiment and selections.

Before you begin

Make sure you read and understand the information in this chapter and have the necessary computer equipment, software, experiment design, and data before you start your analysis.



** **Acquire data** is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Introduction 6
Required items 8
Compliance 10

Agilent Technologies

Introduction

An Integrated Biology (IB) workflow typically combines the results generated by multi-omics analyses into a new experiment. The aim is to find important correlations and validation through statistical analysis, ultimately leading to further insight into a biological system.

This Workflow Guide is complementary to the *Agilent Metabolomics Workflow - Discovery Workflow Guide* (Agilent publication 5990-7067EN) and covers advanced operations available in Mass Profiler Professional (MPP) that help you perform integrated, pathway-level analysis of the primary data from any Agilent omics platform, while also enabling incorporation of prior knowledge - existing datasets, pathway maps, and interaction maps - for greater analytical power in your multi-omics experiments.

Metabolomic studies involve the process of identification and quantification of the endogenous components that form a chemical fingerprint of an organism, or situation under study, and may involve the process of identifying correlations related to changes in the fingerprint as affected by external parameters (metabonomics). Mass Profiler Professional may be used in the study of metabolomics and metabonomics for small molecule studies, proteomics for protein biomarker studies, and general differential analysis. Regardless of the specific study and molecular class, the process is referred to as "metabolomics" throughout this workflow.

To increase your confidence in obtaining reliable and statistically significant results, review the chapter "Prepare for an experiment" in the *Agilent Metabolomics Workflow - Discovery Workflow Guide* and make sure your analysis includes a carefully thought-out experimental design that includes the collection of replicate samples.

More information

The *Integrated Biology with Mass Profiler Professional Workflow Guide* is part of the collection of Agilent manuals, help, application notes, and training videos. The current collection of manuals and help are valuable to users who understand the metabolomics workflow and who may require familiarization with the Agilent software tools. Training videos provide step-by-step instructions for using the software tools to reduce example GC/MS and LC/MS data but require a significant time investment and ability to extrapolate the example processes. This workflow provides a step-by-step overview of performing metabolomics data analysis using Agilent MassHunter Qualitative Analysis and Agilent Mass Profiler Professional.

The following selection of publications provides materials related to metabolomics and Agilent MassHunter Mass Profiler Professional software:

- *Manual:* Agilent Metabolomics Workflow - Discovery Workflow Guide (5990-706EN, Revision B, October 2012)
- *Manual:* Agilent Metabolomics Workflow - Discovery Workflow Overview (5990-7069EN, Revision B, October 2012)
- *Manual:* Agilent G3835AA MassHunter Mass Profiler Professional - Quick Start Guide (G3835-90009, Revision A, November 2012)
- *Manual:* Agilent G3835AA MassHunter Mass Profiler Professional - Familiarization Guide (G3835-90010, Revision A, November 2012)
- *Manual:* Agilent G3835AA MassHunter Mass Profiler Professional - Application Guide (G3835-90011, Revision A, November 2012)

- *Presentation:* Advances in Instrumentation and Software for Metabolomics Research (Advances in Instrumentation and Software for Metabolomics.pdf, September 18, 2012)
- *Brochure:* Agilent Solutions for Metabolomics (5990-6048EN, April 30, 2012)
- *Brochure:* Agilent Mass Profiler Professional Software (5990-4164EN, April 27, 2012)
- *Application:* Mass Profiler Professional and Personal Compound Database and Library Software Facilitate Compound Identification for Profiling of the Yeast Metabolome (5990-9858EN, April 25, 2012)
- *Brochure:* Pathways to Insight - Integrated Biology at Agilent (5991-0222EN, March 30, 2012)

A complete list of references may be found in "References" on page 120.

NOTE

This manual gives links to most references. If you have an electronic copy of this manual, you can easily download the documents from the Agilent literature library. Look for and click the blue hypertext; for example, you can click the "Agilent literature library" link in the previous sentence.

If you have a printed copy, go to the Agilent literature library at www.agilent.com/chem/library and type the publication number in the **Keywords** or **Part Number** box. Then click **Search**. (Note: If you type the publication number into the **Keywords** box, you find the publication number *and* additional publications that reference the publication number.)

"Definitions" on page 110 contains a list of terms and their definitions as used in this workflow.

Required items

- Agilent MassHunter Data Acquisition Software
- Data from an Agilent mass spectrometer
- PC running Windows
- Agilent Mass Profiler Professional Software
- Agilent MassHunter ID Browser
- Agilent MassHunter Qualitative Analysis Software
- Agilent MassHunter DA Reprocessor

The Integrated Biology with Mass Profiler Professional workflow performs best when using the hardware and software described in the "required" sections below. The required hardware and software is used to perform the data acquisition and analysis tasks shown in Figure 1.

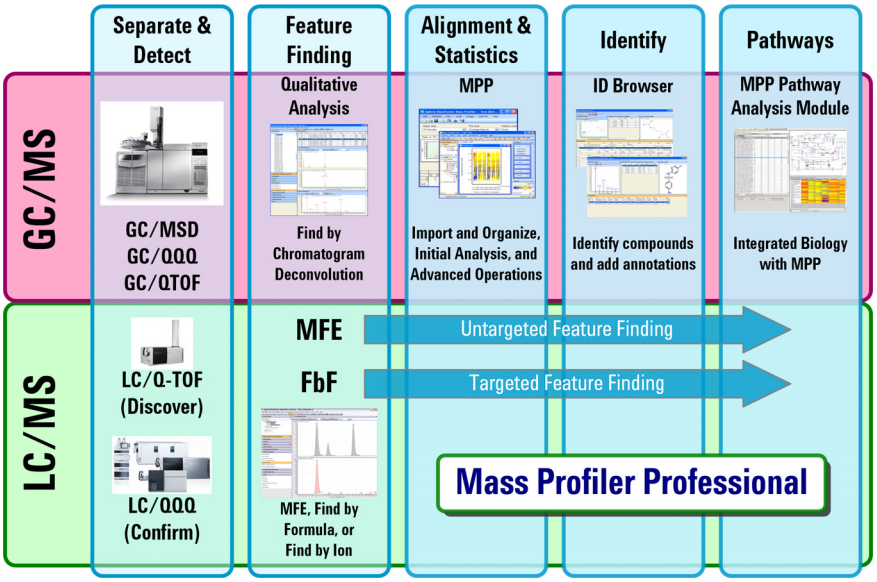


Figure 1 *Agilent hardware and software used to acquire and analyze your samples following the Agilent Integrated Biology Workflow. Sample separation to pathway analysis typically involves either or both GC/MS and LC/MS analyses.*

Required hardware

- PC running Windows
  - *Minimum:* XP SP3 (32-bit) or Windows 7 (32-bit or 64-bit) with 4 GB of RAM
  - *Recommended:* Windows 7 (64-bit) with 8 GB or more of RAM
  - At least 50 GB of free space on the C:\ partition of the hard drive
- Data from an Agilent GC/MS, LC/MS, CE/MS and/or ICP-MS system or data that may be imported from another instrument.

Required software

- Agilent Mass Profiler Professional Software B.12.00 or later
- Agilent MassHunter Qualitative Analysis software, Version B.03.01, B.04.00, B.05.00 SP1 or later
- Agilent MassHunter Data Acquisition software, Version B.03.02, B.04.00, B.05.00 or later (this will include Agilent MassHunter DA Reprocessor)
- Agilent MassHunter Quantitative Analysis software, Version B.03.02 or later

Optional software

- Agilent ChemStation software
- AMDIS
- MassHunter ID Browser B.03.01 or later
- METLIN Personal Compound Database and Library
- Agilent Fiehn GC/MS Metabolomics Library

Compliance

21 CFR Part 11 is a result of the efforts of the US Food and Drug Administration (FDA) and members of the pharmaceutical industry to establish a uniform and enforceable standard by which the FDA considers electronic records equivalent to paper records and electronic signatures equivalent to traditional handwritten signatures. For more information, see

http://www.fda.gov/RegulatoryInformation/Guidances/ucm125067.htm

MassHunter Data Acquisition Compliance Software includes the following features which support 21 CFR Part 11 compliance:

- Hash Signature for data files let you check the integrity of files during a compliance audit
- Roles that restrict actions to certain users
- Method Audit Trail Viewer

MassHunter Quantitative Analysis Compliance Software includes the following features which support 21 CFR Part 11 compliance:

- Security measures ensuring the integrity of acquired data, analysis, and report results
- Comprehensive audit-trail features for quantitative analysis, using a flexible and configurable audit-trail map
- Customizable user roles and groups let an administrator individualize user access to processing tasks

Before you begin creating methods and submitting studies, you may decide to install MassHunter Data Acquisition Compliance Software and MassHunter Quantitative Analysis Compliance Software.

The Quantitative Analysis Compliance program is installed separately from the Quantitative Analysis program. See *Agilent MassHunter Quantitative Analysis Compliance Software Quick Start Guide* (Agilent publication G3335-90099, Revision A, February 2011) for instructions on installing the Compliance program.

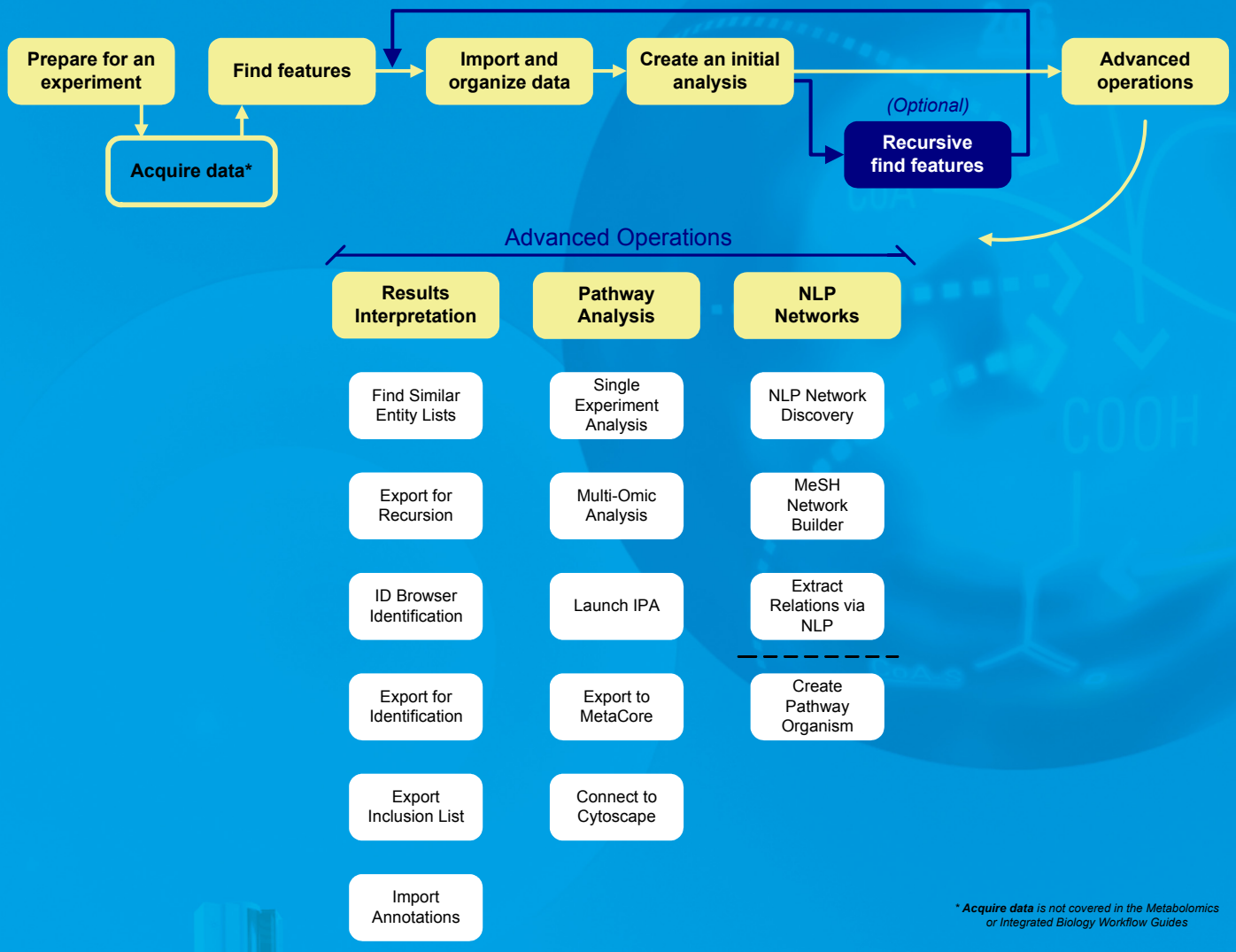
The Data Acquisition Compliance program is installed automatically with the MassHunter Data Acquisition software. See *Agilent MassHunter Data Acquisition Compliance Software Quick Start Guide* (Agilent publication G3335-90098, Revision A, February 2011) for instructions on enabling and using the MassHunter Compliance Software.

Roles

When Compliance is enabled, only certain users can perform certain actions. For example, the user that logs on to the system to submit a study needs to have certain Quantitative Analysis privileges to automatically build the quantitative analysis method.

Working with Mass Profiler Professional

This chapter helps you understand where Mass Profiler Professional is used in a typical metabolomics analysis and directs you to additional documentation that covers using Mass Profiler Professional.



* ***Acquire data*** *is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Where is MPP used in your experiment? 12
What is the metabolomics workflow? 13
Advanced operations covered in the MPP workflow guides 16
Using Mass Profiler Professional 17

Agilent Technologies

Where is MPP used in your experiment?

Mass Profiler Professional is used to import, organize, and analyze the data you acquired from your experimental samples. Your untargeted differential analysis experiment may include eight steps as shown below. Mass Profiler Professional begins at step four.

(1) Prepare for your experiment
(2) Acquire your data
(3) Find the spectral features
(4) Import and organize your data
(5) Create your initial analysis
(6) Identify the features
(7) Save your project
(8) Perform advanced analysis operations

Figure 2 shows the steps and Agilent tools that are used in your experiment.

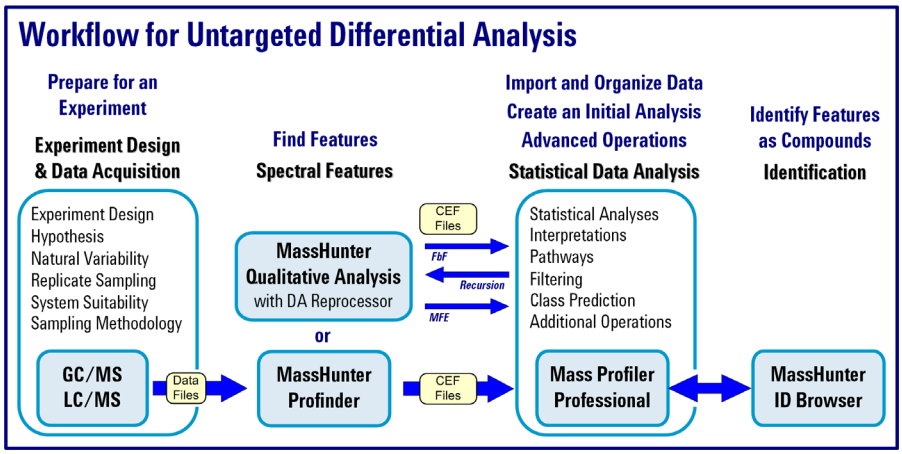


Figure 2 *The steps involved in an untargeted differential analysis.*

What is the metabolomics workflow?

Metabolomics is an emerging field of 'omics' research that is concerned with the characterization and identification of the metabolite content of a cell or whole organism. Metabolomics studies let researchers view biological systems in a way that is different from but complementary to genomics, transcriptomics, and proteomics studies. Discovery metabolomics experiments involve examining an untargeted suite of metabolites, finding the metabolites with statistically significant variations in abundance within a set of experimental versus control samples, and answering questions related to causality and relationships. Metabolomics is a powerful, emerging discipline with a broad range of applications, including basic research, clinical research, drug development, environmental toxicology, crop optimization, and food science.

Metabolomics research leads to complex data sets involving hundreds to thousands of metabolites. Comprehensive analysis of metabolomics data requires an analytical approach and data analysis strategy that are often unique and require specialized data analysis software that enables cheminformatics analysis, bioinformatics, and statistics. Agilent provides you with tools to perform metabolomics research.

Experiment variables are derived from your experiment. When one or more of the attributes of the state of the organism are manipulated those attributes are referred to as independent variables. The biological response to the change in the attributes may manifest in a change in the metabolic profile. Each metabolite that undergoes a change in expressed concentration is referred to as a dependent variable. Metabolites that do not show any change with respect to the independent variable may be valuable as control or reference signals.

The metabolites in a sample may be individually referred to as a **compound**, **feature**, **element**, or **entity** during the various steps of the metabolomic data analysis. When hundreds to thousands of dependent variables (e.g., metabolites) are available, chemometric data analyses is employed to reveal accurate and statistically meaningful correlations between the attributes (independent variables) and the metabolic profile (dependent variables). Meaningful information learned from the metabolite responses can be part of a larger process that is used to develop clinical diagnostics, for understanding the onset and progression of human diseases, and for treatment assessment. Therefore, metabolomic analyses are poised to answer questions related to causality and relationship as applied to chemically complex systems, such as organisms.

You can use a metabolomics workflow as a road map for any analysis that requires the identification of statistically significant answers to questions presented to complex data sets. The metabolomics workflow may be used to perform the following analyses:

- Compare two or more biological groups
- Find and identify potential biomarkers
- Look for biomarkers of toxicology
- Understand biological pathways
- Discover new metabolites
- Develop data mining and data processing procedures that produce characteristic markers for a set of samples
- Construct statistical models for sample classification.

Typical metabolomics workflow

A typical Agilent metabolomics workflow is illustrated in Figure 3 starting with data acquisition through to analysis involving both untargeted (discovery) LC/MS and targeted (confirmation) LC/MS/MS analyses. Molecular feature extraction (MFE) and Find by Formula (FbF) are two different algorithms used by MassHunter Qualitative Analysis for finding compounds. All results files generated by Agilent analytical platforms can be imported into Mass Profiler Professional for quality control, statistical analysis, visualization, and interpretation.

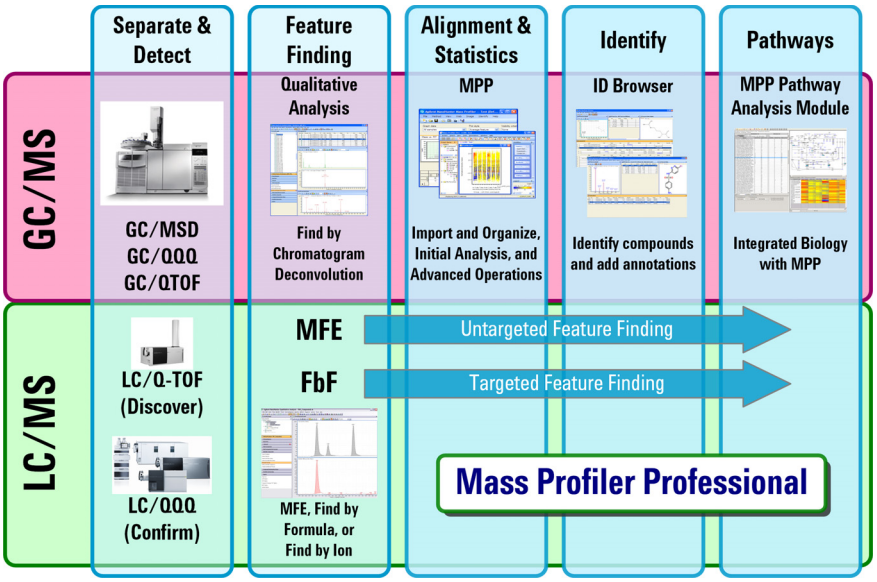


Figure 3 *An Agilent metabolomics workflow from separation to pathway analysis typically involves either or both GC/MS and LC/MS analyses.*

Variables

A metabolomics workflow analysis involves two types of variables that are associated with your samples:

Independent variables: One or more of the attributes of the state of the organism that are known to you in advance of sampling. These attributes are referred to as an independent variable.

During the various steps of the data analysis the workflow refers to the known states of the organism, or externalities to which the organism is subjected, as parameter values, conditions, or attribute values. The known states and externalities represent independent variables in the statistical analyses.

Dependent variables: The observable biological response to changes in the independent variables. The response can manifest as a change in the metabolic profile. Each metabolite that undergoes a change in expressed concentration is referred to as a dependent variable.

The metabolites in a sample may be individually referred to as compounds, features, elements, or entities during the various steps of the metabolomic data analysis. Metabolites represent dependent variables in the statistical analyses.

The hypothesis

The first and most important step in your experiment is to formulate the question of correlation that is answered by the analysis - the hypothesis. This question is a statement that proposes a possible correlation, for example a cause and effect, between a set of independent variables and the resulting metabolic profile. The workflow is used to prove or disprove the hypothesis.

Natural variability

Before your begin collecting your samples it is important to understand how any one sample represents the population as a whole. Because of natural variability and the uncertainties associated with both the measurement and the population, no assurance exists that any single sample from a population represents the mean of the population. Thus, increasing the sample size greatly improves the accuracy of the sample set in describing the characteristics of the population.

Replicate sampling

Sampling the entire population is not typically feasible because of constraints imposed by time, resources, and finances. On the other hand, fewer samples increase the probability of concluding a false positive or false negative correlation. At a minimum, it is recommended that your analysis include ten (10) or more replicate samples for each attribute value for each condition in your study.

System suitability

System suitability involves collecting data to provide you with a means to evaluate and compensate for drift and instrumental variations to assure quality results. The techniques that produce the highest quality results include (1) retention time alignment, (2) intensity normalization, (3) chromatographic deconvolution, and (4) baselining. However, even the best analysis techniques cannot compensate for excessive drift in the acquisition parameters. The best results are achieved by maintaining your instrument and using good chromatography.

Sampling methodology

Improved data quality for your analysis comes from matching the sampling methodology to the experimental design so that replicate data is collected to span the attribute values for each condition. A larger number of samples appropriate to the population under study results in a better answer to the hypothesis. An understanding of the methodologies used in sampling and using more than one method of sample collection have a positive impact on the significance of your results.

Advanced operations covered in the MPP workflow guides

In many cases the example data used in this workflow is processed using the metabolomics workflow before being analyzed using the integrated biology operations. Familiarity with the terminology and steps described in the *Agilent Metabolomics Workflow - Discovery Workflow Guide* with help you use this workflow guide and the advanced operations used in integrated biology. Figure 4 shows a summary of the Metabolomics Discovery Workflow and the advanced operations covered by the both the metabolomics and the integrated biology workflow guides.

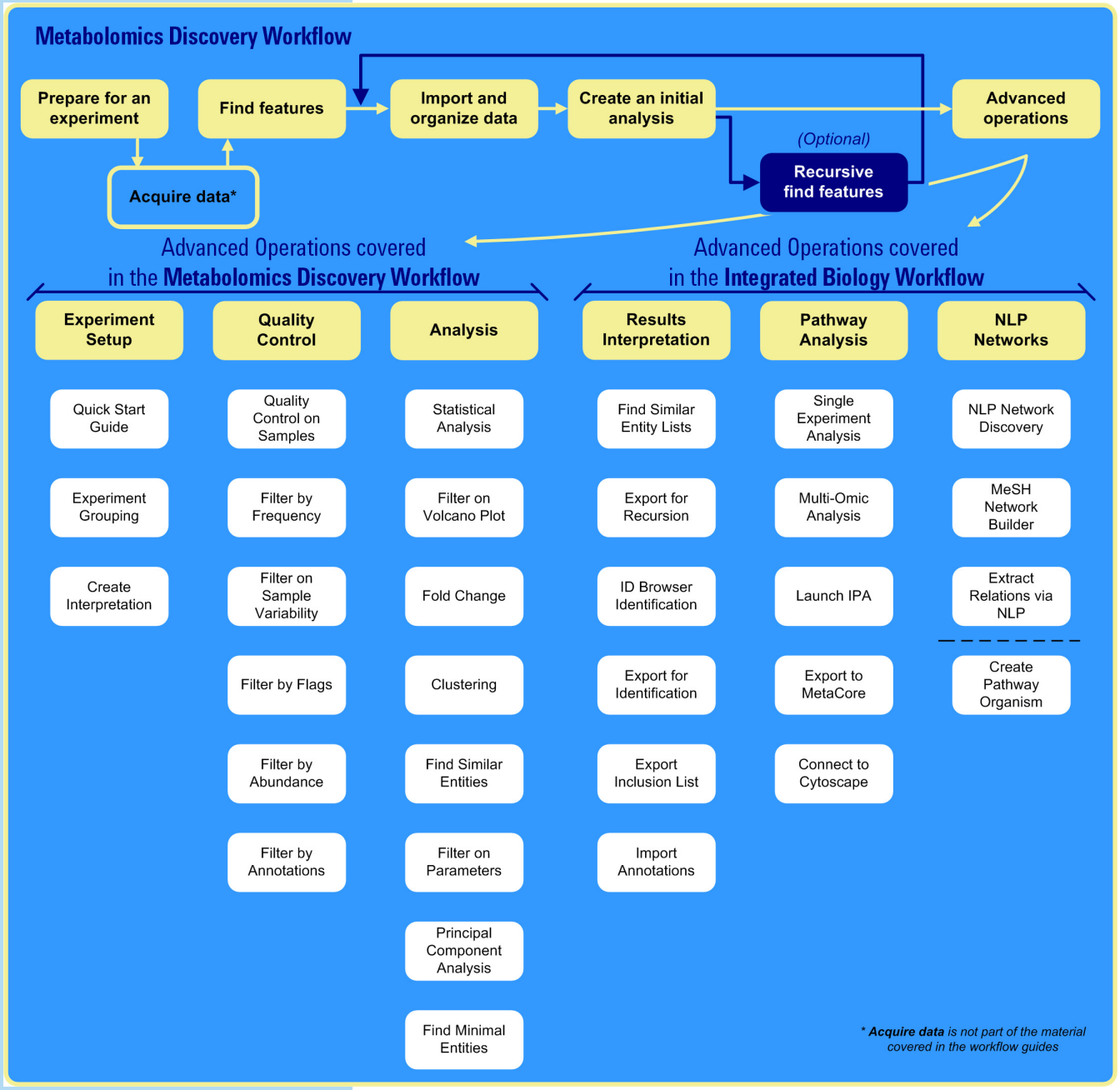


Figure 4 *Summary of the Metabolomics Discovery Workflow and MPP advanced operations covered in the Metabolomics and Integrated Biology workflows.*

Using Mass Profiler Professional

Mass Profiler Professional helps you analyze your data through the use of sequential dialog boxes and wizards as shown in Figure 5.

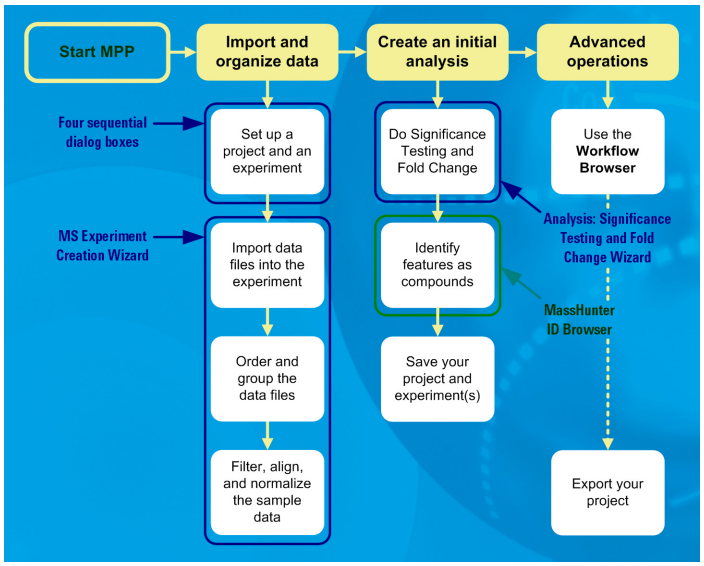

Start MPP
Import and organize data
Create an initial analysis
Advanced operations
Four sequential dialog boxes
Set up a project and an experiment
Do Significance Testing and Fold Change
Use the Workflow Browser
MS Experiment Creation Wizard
Import data files into the experiment
Identify features as compounds
Analysis: Significance Testing and Fold Change Wizard
MassHunter ID Browser
Order and group the data files
Save your project and experiment(s)
Filter, align, and normalize the sample data
Export your project

Figure 5 *Overview of the wizards that help you use Mass Profiler Professional.*

A series of guides are available from the Agilent Literature Library (http://www.chem.agilent.com/en-US/Search/Library/Pages/default.aspx) to help you become familiar with using Mass Profiler Professional and preparing for your experiment.

The *Agilent G3835AA MassHunter Mass Profiler Professional - Quick Start Guide* (Agilent publication G3835-90009) helps you launch MPP, activate your license, review the MPP user interface, and create a project and an experiment that you import preloaded data into and then use to begin a sample analysis.

The *Agilent G3835AA MassHunter Mass Profiler Professional - Familiarization Guide* (Agilent publication G3835-90010) provides a familiarization tutorial that helps you create your first project and experiment using MPP.

The *Agilent G3835AA MassHunter Mass Profiler Professional - Application Guide* (Agilent publication G3835-90011) helps you prepare for your experiment and guide you through an untargeted differential analysis of your data.

The *Agilent Metabolomics Workflow - Discovery Workflow Guide* (Agilent publication 5990-7067EN) provides you with additional detail, techniques, and explanations to improve your experiment design and perform advanced analysis operations.

Layout of the Mass Profiler Professional screen

The main functional areas of the Mass Profiler Professional screen are illustrated in Figure 6.

The main Mass Profiler Professional window consists of four parts:

Menu Bar - access to actions that are used for managing your projects, experiments, pathways, and display pane views
Toolbar - access to buttons for commonly used tasks grouped by project, experiment, entity, statistical plot, and sidebar tasks
Display Pane - organized into functional areas that help you navigate through your project, experiments, analyses, and available operations
Status Bar - information related to the current view, cursor position, entity, and system memory

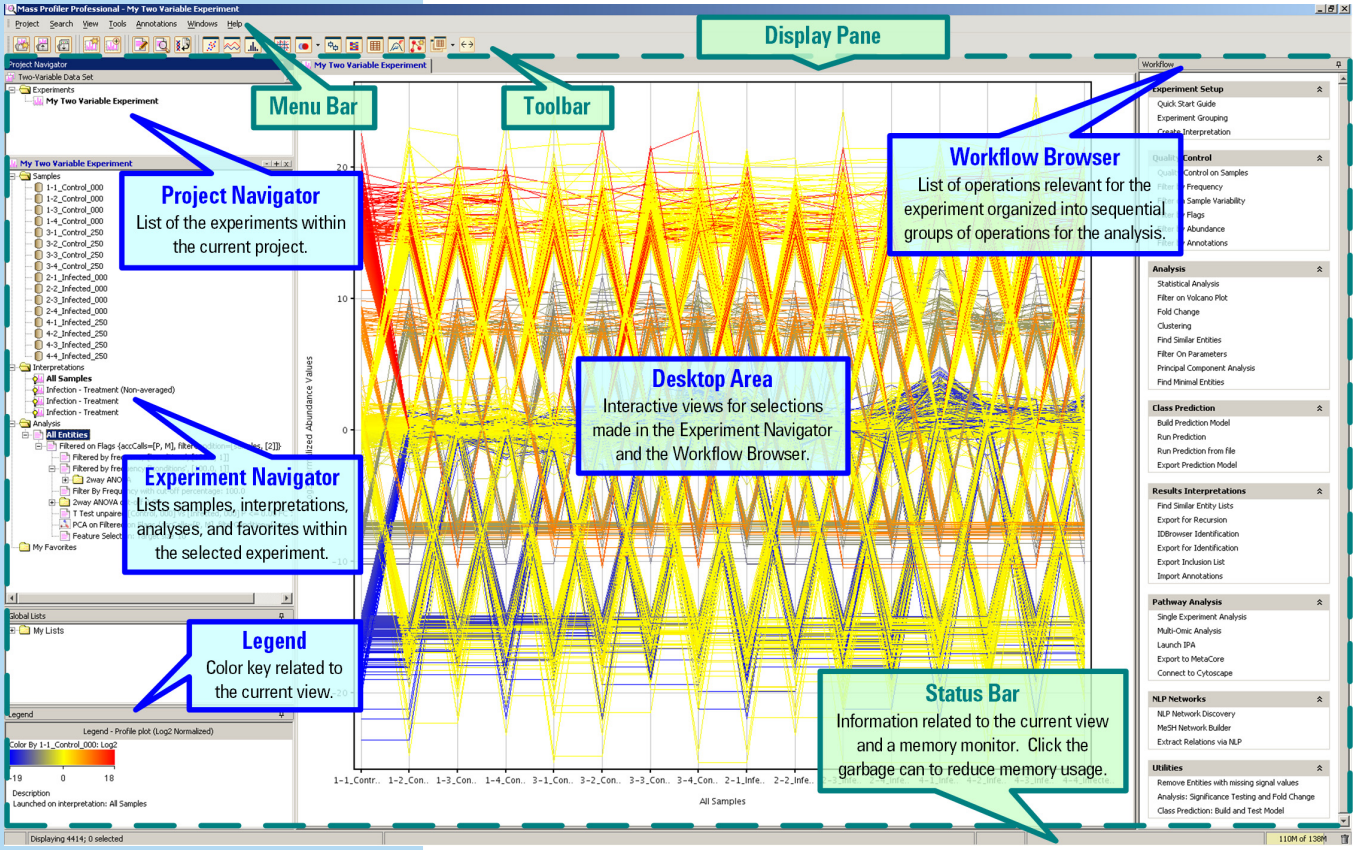
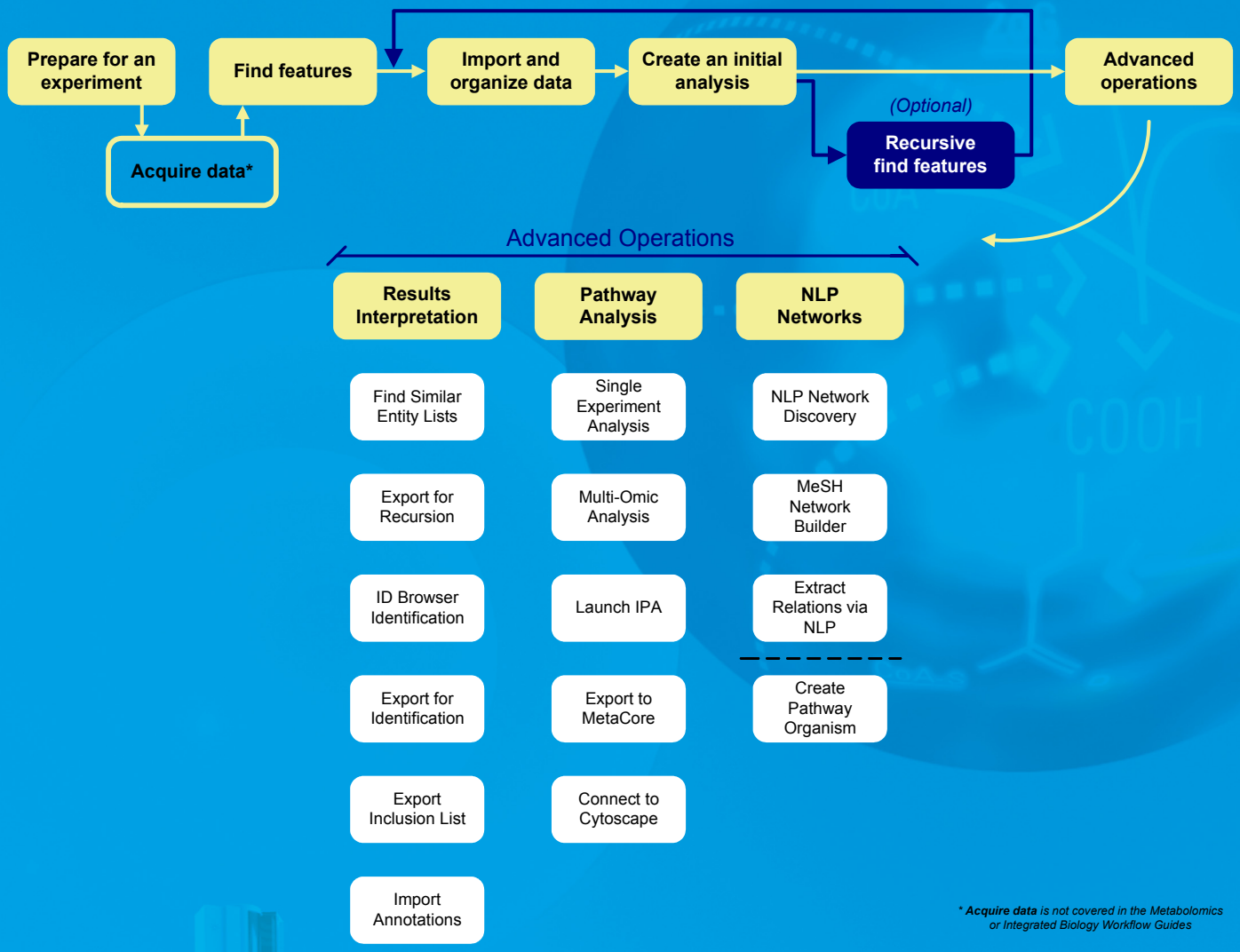

Display Pane
Menu Bar
Toolbar
Project Navigator
List of the experiments within the current project.
Workflow Browser
List of operations relevant for the experiment organized into sequential groups of operations for the analysis.
Desktop Area
Interactive views for selections made in the Experiment Navigator and the Workflow Browser.
Experiment Navigator
Lists samples, interpretations, analyses, and favorites within the selected experiment.
Legend
Color key related to the current view.
Status Bar
Information related to the current view and a memory monitor. Click the garbage can to reduce memory usage.

Figure 6 *The main functional areas of Mass Profiler Professional*

Example experiments

The experiments described in this chapter allow the workflow to guide you through the options available for your analysis.



* ***Acquire data*** *is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Features of the example mass spectrometry experiments 20
Features of the example array experiment 22
Creating an expression analysis using the sample array experiment 23

Features of the example mass spectrometry experiments

The mass spectrometry analysis capabilities of Mass Profiler Professional are illustrated in this workflow with experiments that contain either (1) a single independent variable or (2) two independent variables. Each of the advanced operations available in the Workflow Browser use a wizard to guide you through the operation. The steps and wizard pages may change each time you perform the operation depending on the number of variables in your experiment and analysis features selected. The two experiments described below allow this workflow to guide you through the options available for your analysis.

Definitions

Terms and definitions used in metabolomics and metabolomic analyses vary. It is recommended that you refer to the "Definitions" on page 110 for a list of terms and their definitions as used in Mass Profiler Professional and in this workflow.

One-variable experiment

The one-variable experiment presents an analysis of a metabolomic response to changes in a single independent variable, also referred to as a parameter. The data was acquired using four (4) parameter values for the independent variable. The parameter values consist of a single control data set that represents the organism without perturbation and data sets from three variations where the organism is subject to one of three conditions established by the experiment design. In summary, the one-variable experiment contains a single parameter with four parameter values and ten replicate samples for each parameter value.

Based on the discussion presented in the "Prepare for an experiment" chapter in the *Agilent Metabolomics Workflow - Discovery Workflow Guide*, an ideal experiment involves at least ten (10) replicates for each parameter value. Thus an ideal experiment with a single parameter and four parameter values has a data sample size of at least forty (40) samples. In this example the minimum sampling conditions are met.

In the experiment sample list shown in Figure 7 the parameter values for the independent variable are listed in the Group ID column. Since sample names are derived from your actual data file names, CEF files in this example, it is recommended to develop a concise, meaningful file naming convention for your experiment.

Samples	Group ID
01-Control	Control
02-Control	Control
03-Control	Control
04-Control	Control
05-Control	Control
06-Control	Control
07-Control	Control
08-Control	Control
09-Control	Control
10-Control	Control
11-Variation AW	Set A
12-Variation AW	Set A
13-Variation AW	Set A
14-Variation AW	Set A
15-Variation AW	Set A
16-Variation AW	Set A
17-Variation AW	Set A
18-Variation AW	Set A
19-Variation AW	Set A
20-Variation AW	Set A

Samples	Group ID
21-Variation BR	Set B
22-Variation BR	Set B
23-Variation BR	Set B
24-Variation BR	Set B
25-Variation BR	Set B
26-Variation BR	Set B
27-Variation BR	Set B
28-Variation BR	Set B
29-Variation BR	Set B
30-Variation BR	Set B
31-Variation CN	Set C
32-Variation CN	Set C
33-Variation CN	Set C
34-Variation CN	Set C
35-Variation CN	Set C
36-Variation CN	Set C
37-Variation CN	Set C
38-Variation CN	Set C
39-Variation CN	Set C
40-Variation CN	Set C

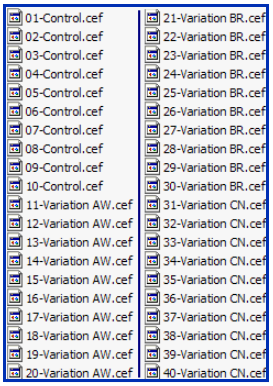


Figure 7 *One-variable experiment sample list and file list*

Two-variable experiment

The two-variable experiment presents an analysis of a metabolomic response to changes in two independent variables (parameters), each with two parameter values. The parameter values of the first parameter represent a control data set associated with the organism without perturbation and when the organism was subject to a known perturbation. The parameter values of the second parameter represent a pair of metabolite extraction techniques where the first parameter value represents the current state-of-the-art extraction process and the second parameter value represent the addition of a step designed to improve metabolite extraction. In summary, the two-variable experiment contains two parameters with two parameter values, for a total of four permutations, and four replicate samples were obtained for each permutation.

Based on the discussion presented in the "Prepare for an experiment" chapter in the *Agilent Metabolomics Workflow - Discovery Workflow Guide*, an ideal experiment involves at least ten (10) replicates for each parameter value. Thus an ideal experiment with two parameters, each with two parameter values, has a data sample size of forty (40) samples. The ideal sample size is calculated by multiplying 2 parameters by 2 parameter values for each parameter and then multiplying by 10 replicates for an ideal minimum sample size of forty (2 x 2 x 10 = 40) samples. In this example the minimum sampling conditions are not met; four replicates exist for each permutation for a total of sixteen (16) samples. While the sampling falls short of the minimum sampling recommendation, the strong correlation of cause and effect in this experiment overcomes the sampling deficiency and provides support for further investment in the metabolomics question being studied.

In the experiment sample list shown in Figure 8 the parameter values for the independent variables are listed in the Infection and Treatment columns. Since sample names are derived from your actual data file names, CEF files in this example, it is recommended to develop a concise, meaningful file naming convention for your experiment.

Samples	Infection	Treatment
1-1_Control_000	Control	0
1-2_Control_000	Control	0
1-3_Control_000	Control	0
1-4_Control_000	Control	0
2-1_Infected_000	Infected	0
2-2_Infected_000	Infected	0
2-3_Infected_000	Infected	0
2-4_Infected_000	Infected	0
3-1_Control_250	Control	250
3-2_Control_250	Control	250
3-3_Control_250	Control	250
3-4_Control_250	Control	250
4-1_Infected_250	Infected	250
4-2_Infected_250	Infected	250
4-3_Infected_250	Infected	250
4-4_Infected_250	Infected	250

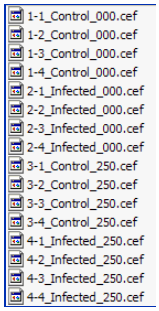


Figure 8 *Two-variable experiment sample list and file list*

Features of the example array experiment

Some pathway analysis capabilities of Mass Profiler Professional are illustrated in this workflow with an array experiments that contains a single independent variable. Each of the advanced operations available in the Workflow Browser use a wizard to guide you through the operation. The steps and wizard pages may change each time you perform the operation depending on the number of variables in your experiment and analysis features selected. The experiment described below allow this workflow to guide you through the options available for your analysis.

Definitions

Terms and definitions used in metabolomics and metabolomic analyses vary. It is recommended that you refer to the "Definitions" on page 110 for a list of terms and their definitions as used in Mass Profiler Professional and in this workflow.

One-variable array experiment

The one-variable array experiment presents an analysis of a treated versus untreated sample; changes in a single independent variable, also referred to as a parameter. The data was acquired using two (2) parameter values for the independent variable. The parameter values consist of a single control data set that represents the sample without perturbation and a data set from a variation where the sample was treated to a conditions established by the experiment design. In summary, the one-variable experiment contains a single parameter with two parameter values and three replicate samples for each parameter value.

Based on the discussion presented in the "Prepare for an experiment" chapter in the *Agilent Metabolomics Workflow - Discovery Workflow Guide*, an ideal experiment involves at least ten (10) replicates for each parameter value. Thus an ideal experiment with a single parameter and two parameter values has a data sample size of at least twenty (2 x 10 = 20) samples. In this example the minimum sampling conditions are not met. three replicates exist for each permutation for a total of six (6) samples. While the sampling falls short of the minimum sampling recommendation, the strong correlation of cause and effect in this experiment overcomes the sampling deficiency and provides support for further investment in the question being studied.

In the experiment sample list shown in Figure 9 the parameter values for the independent variable are listed in the Treatment column. Since sample names are derived from your actual data file names, text files in this example, it is recommended to develop a concise, meaningful file naming convention for your experiment.

Samples	Treatment
US22502705_251209747382_Untreated.txt	Untreated
US22502705_251209747387_Untreated.txt	Untreated
US22502705_251209747392_Treated.txt	Treated
US22502705_251209747393_Treated.txt	Treated
US22502705_251209747394_Untreated.txt	Untreated
US22502705_251209747404_Treated.txt	Treated

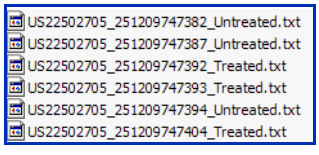


Figure 9 *One-variable array experiment sample list and file list*

Creating an expression analysis using the sample array experiment

The workflow for importing and performing an initial analysis of gene probe data is different from the workflow used for mass spectral data as described in the *Agilent Metabolomics Workflow - Discovery Workflow Guide*. This section guides you through steps necessary to import and prepare the **Agilent Expression Single Color Demo** sample data installed with Mass Profiler Professional to demonstrate some of the advanced operations in this workflow.

MPP is used to import, organize, and analyze the data you acquired. An experiment based on **Expression** selected for the **Analysis type** using the **Agilent Expression Single Color Demo** sample data includes the following steps: (1) create a project and experiment, (2) import your data, (3) create your initial analysis, and (4) perform advanced analysis operations. Figure 10 shows these steps to prepare the **Agilent Expression Single Color Demo** sample data to become familiar with the integrated biology operations.

The **Analysis: Biological Significance** wizard guides you through eight (8) steps to organize and enter parameters and values that improve the quality of your results and produce an initial differential expression of the sample data. The steps performed during the **Analysis: Biological Significance** wizard are illustrated in Figure 11 on page 24

The entity list created from the **Agilent Expression Single Color Demo** sample data is a gene probe-based entity list rather than a compound-based entity list created from mass spectrometry data.

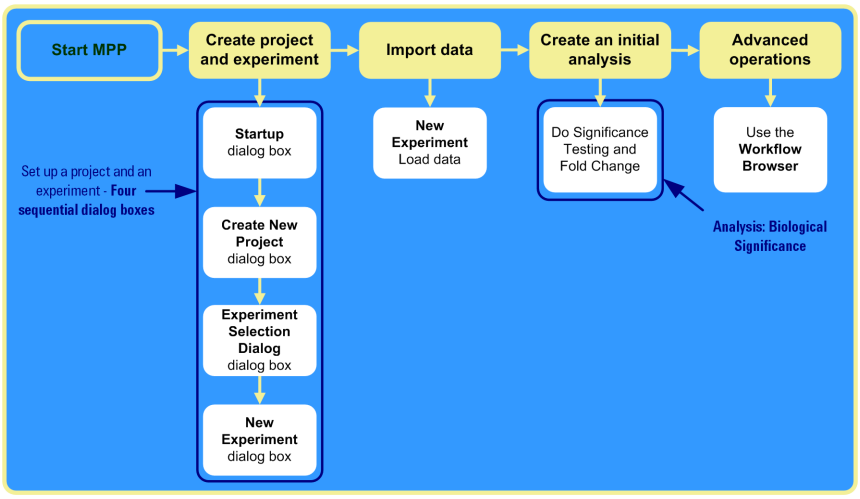

Start MPP
Create project and experiment
Import data
Create an initial analysis
Advanced operations
Startup dialog box
Create New Project dialog box
Experiment Selection Dialog dialog box
New Experiment dialog box
New Experiment Load data
Do Significance Testing and Fold Change
Use the Workflow Browser
Set up a project and an experiment - Four sequential dialog boxes
Analysis: Biological Significance

Figure 10 *The steps to import and analyze the Agilent Expression Single Color Demo sample data.*

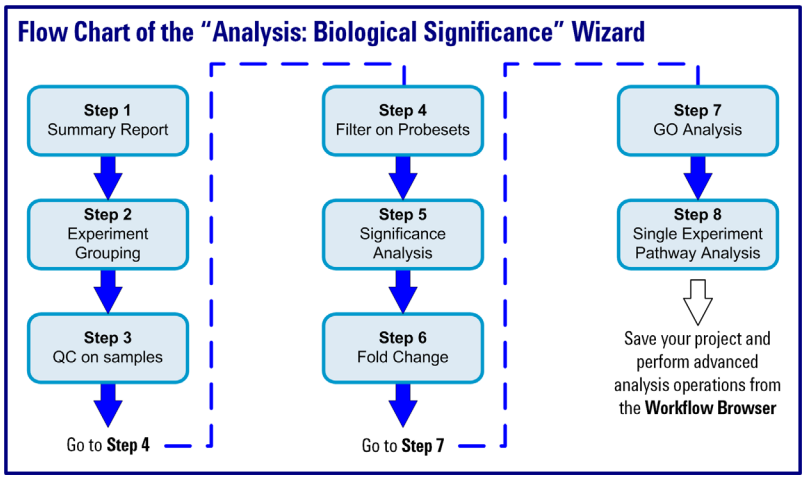


Figure 11 *Steps performed by the Analysis: Biological Significance wizard*

Set up a project and an experiment

A project is a container for a collection of experiments. A project can have multiple experiments on different sample types and organisms. You are guided through four steps to create a new project and experiment to receive your imported data:

- **Startup:** Select creation of a new project.
- **Create New Project:** Type descriptive information about the project.
- **Experiment Selection Dialog:** Select create a new experiment as part of the project.
- **New Experiment:** Type and select custom information to store with the experiment.

1. Create a new project in the **Startup** dialog box.

a Click **Create new project**.

b Click **OK**.

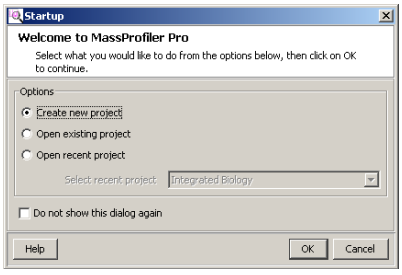


Figure 12 *Welcome to Mass Profiler Professional startup dialog box*

2. Enter descriptive information in the **Create New Project** dialog box.

a Type a descriptive **Name** for the project, `Agilent Single Color Demo`.

b Type descriptive **Notes** for the project.

c Click **OK**.

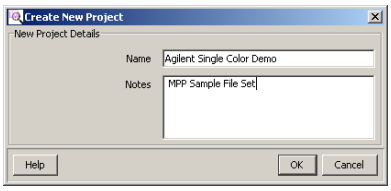


Figure 13 *Create New Project dialog box*

3. Select your experiment origin in the **Experiment Selection Dialog** dialog box.

Specify whether the wizard guides you through creating a new experiment or whether the wizard opens an existing experiment.

a Click **Create new experiment**.

b Click **OK**.

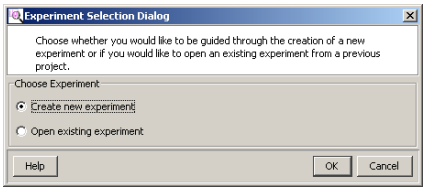

Experiment Selection Dialog
Choose whether you would like to be guided through the creation of a new experiment or if you would like to open an existing experiment from a previous project.
Choose Experiment
Create new experiment
Open existing experiment
Help
OK
Cancel

Figure 14 *Experiment Selection Dialog dialog box*

4. Type and select information that guides the experiment creation in the **New Experiment** dialog box.

Available entry options for the New Experiment dialog box depend on your experiment type and data sources.

a Type a descriptive name for the experiment in **Experiment name**, `Agilent Single Color Demo`.

b Select **Expression** for the **Analysis type**. Only your licensed analysis types are available.

c Select **Agilent Expression Single Color** for the **Experiment type**.

d Select **Analysis: Biological Significance** for the **Workflow type**.

e Type descriptive notes for the experiment in the **Experiment notes**.

f Click **OK**.

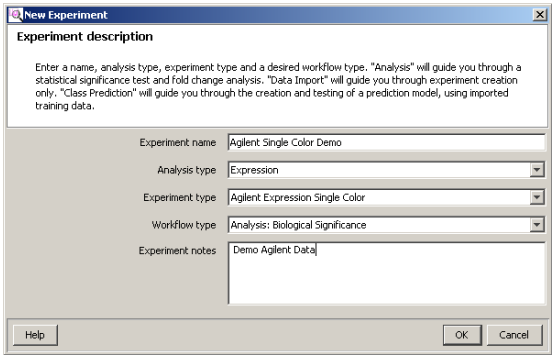

New Experiment
Experiment description
Enter a name, analysis type, experiment type and a desired workflow type. "Analysis" will guide you through a statistical significance test and fold change analysis. "Data Import" will guide you through experiment creation only. "Class Prediction" will guide you through the creation and testing of a prediction model, using imported training data.
Experiment name
Agilent Single Color Demo
Analysis type
Expression
Experiment type
Agilent Expression Single Color
Workflow type
Analysis: Biological Significance
Experiment notes
Demo Agilent Data
Help
OK
Cancel

Figure 15 *Experiment description in the New Experiment dialog box*

Import the sample data

1. Load data from the **New Experiment** dialog box.

a Click **Choose Files**.

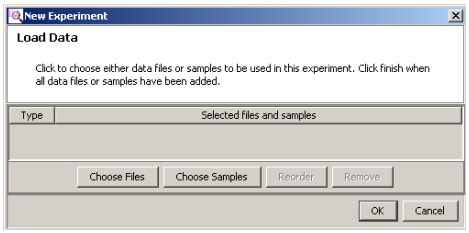


Figure 16 *Load Data from the New Experiment dialog box*

2. Select the sample files in the **Open** dialog box.

a Select the sample data files to open.

b Click **Open**.

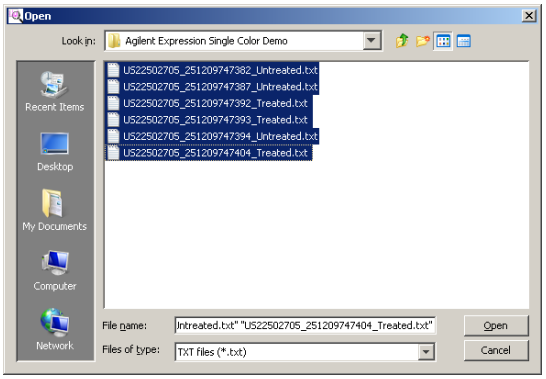


Figure 17 *Open dialog box*

3. Review the sample data in the **New Experiment** dialog box.

a Review the selected sample files.

b Click **OK**. A progress dialog box is shown while importing the sample files.

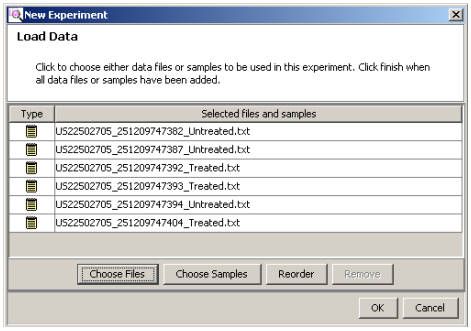


Figure 18 *Experiment Selection Dialog dialog box*

Do Significance Testing and Fold Change

1. Review the summary report in the **Analysis: Biological Significance (Step 1 of 8)** wizard.

The **Analysis: Biological Significance** wizard starts if **Analysis: Biological Significance** was selected as the **Workflow type** in the **New Experiment** dialog box (Figure 15 on page 25).

a Review the data, change the plot view, export selected data, or export the plot to a file, click and right-click features available on the plot.

b Click **Next**.

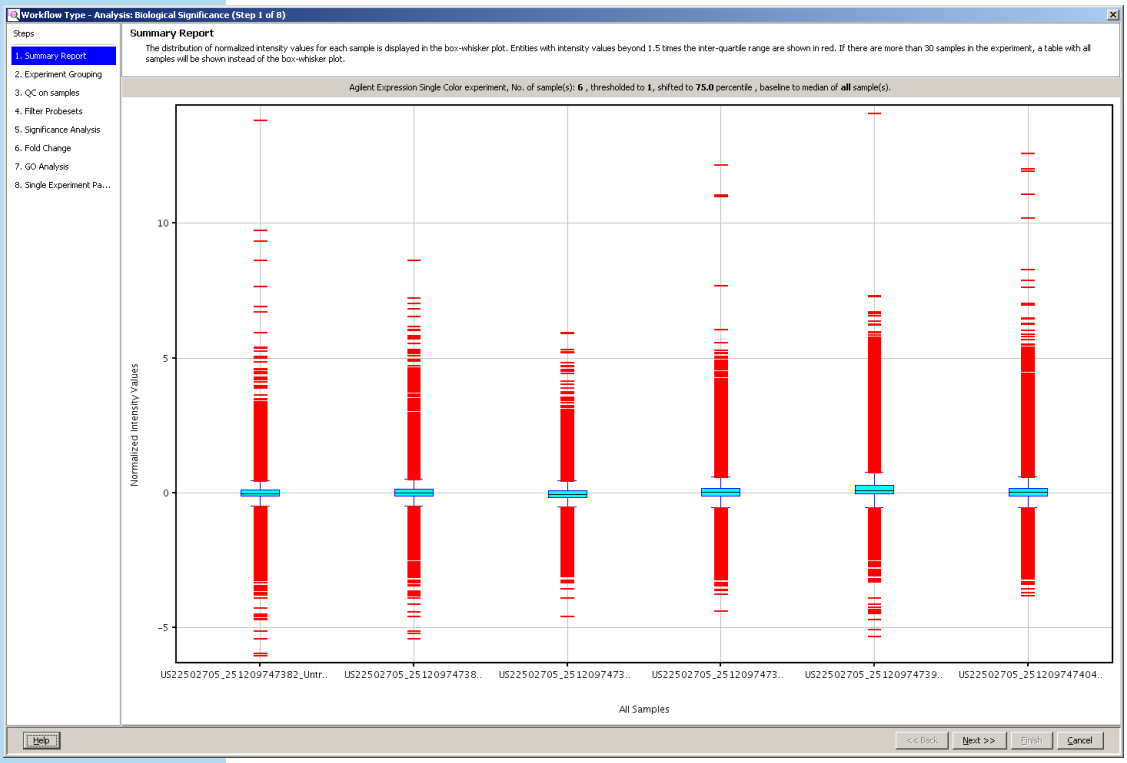

Workflow Type - Analysis: Biological Significance (Step 1 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
Summary Report
The distribution of normalized intensity values for each sample is displayed in the box-whisker plot. Entities with intensity values beyond 1.5 times the inter-quartile range are shown in red. If there are more than 30 samples in the experiment, a table with all samples will be shown instead of the box-whisker plot.
Agilent Expression Single Color experiment, No. of sample(s): 6 , thresholded to 1, shifted to 75.0 percentile , baseline to median of all sample(s).
Normalized Intensity Values
10
5
0
-5
US22502705_251209747382_Untr..
US22502705_25120974738..
US22502705_2512097473..
US22502705_2512097473..
US22502705_25120974739..
US22502705_251209747404..
All Samples
Help
<< Back
Next >>
Finish
Cancel

Figure 19 *Summary Report plot of the Agilent Expression Single Color Demo sample data*

2. Enter the experiment grouping parameters associated with the independent variables and their attribute values in the **Analysis: Biological Significance (Step 2 of 8)** wizard.

In this step you enter your experiment grouping. An independent variable is referred to as a parameter name. The attribute values within an independent variable are referred to as parameter values. Samples with the same parameter values within a parameter name are treated as replicates.

Note: In order to proceed, at least one parameter with two values must be assigned.

Note: When entering Parameter Names and parameter Assign Values, it is very important that the entries use identical letters, numbers, punctuation, and case in order for the Experiment Grouping to function properly. Click **Back** or **Experiment Setup > Experiment Grouping** to return to Experiment Grouping if an error is identified later in the wizard or while performing operations in the Workflow Browser, respectively.

a Click **Add Parameter**.

b Click the **Load experiment parameters from file** button to apply a previously created experiment grouping associated with the sample data.

c Select the file **EXPERIMENT PARAMETERS (can be loaded from file).tsv**.

d Click **Open**. The sample files are automatically grouped and assigned parameter names and parameter values.

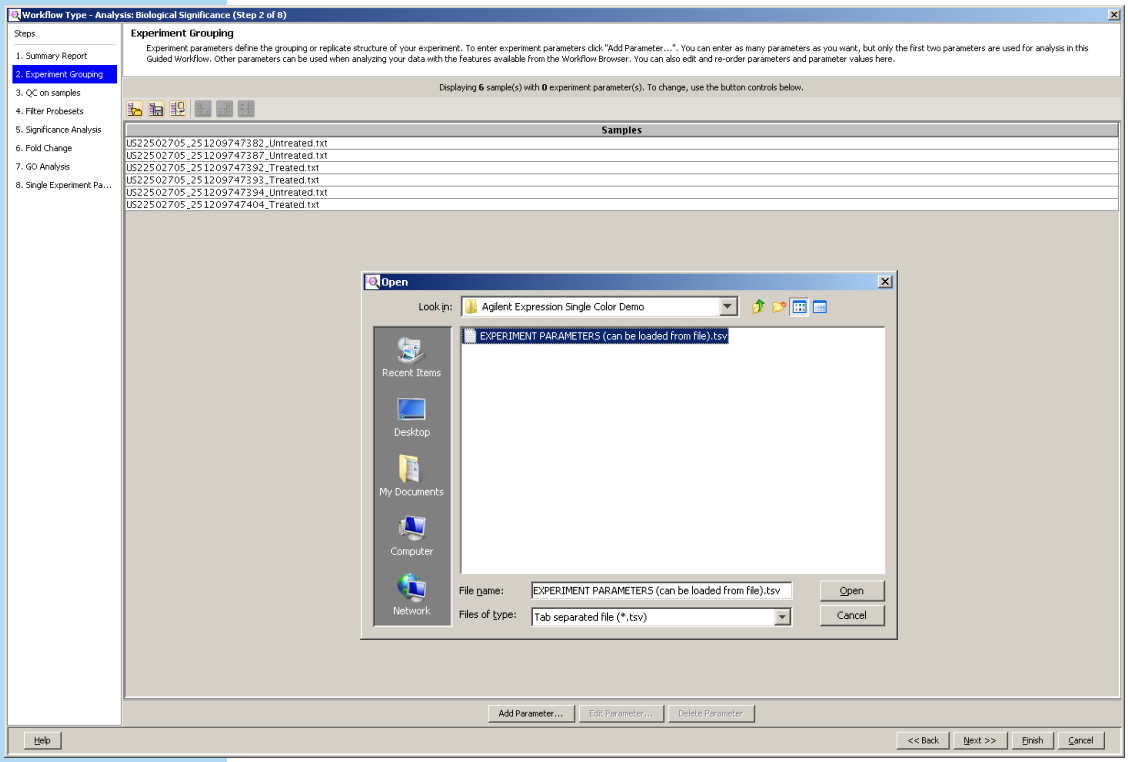

Workflow Type - Analysis: Biological Significance (Step 2 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
Experiment Grouping
Experiment parameters define the grouping or replicate structure of your experiment. To enter experiment parameters click "Add Parameter...". You can enter as many parameters as you want, but only the first two parameters are used for analysis in this Guided Workflow. Other parameters can be used when analyzing your data with the features available from the Workflow Browser. You can also edit and re-order parameters and parameter values here.
Displaying 6 sample(s) with 0 experiment parameter(s). To change, use the button controls below.
Samples
US22502705_251209747382_Untreated.txt
US22502705_251209747387_Untreated.txt
US22502705_251209747392_Treated.txt
US22502705_251209747393_Treated.txt
US22502705_251209747394_Untreated.txt
US22502705_251209747404_Treated.txt
Open
Look in:
Agilent Expression Single Color Demo
EXPERIMENT PARAMETERS (can be loaded from file).tsv
Recent Items
Desktop
My Documents
Computer
Network
File name:
EXPERIMENT PARAMETERS (can be loaded from file).tsv
Open
Files of type:
Tab separated file (*.tsv)
Cancel
Add Parameter...
Edit Parameter...
Delete Parameter
Help
<< Back
Next >>
Finish
Cancel

Figure 20 *Experiment Grouping and loading experiment parameters from file for the Agilent Expression Single Color Demo sample data*

e Click **Next**.

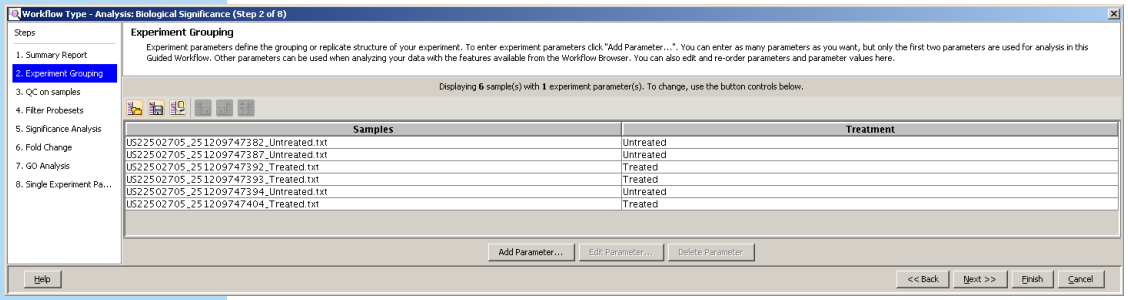

Workflow Type - Analysis: Biological Significance (Step 2 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
Experiment Grouping
Experiment parameters define the grouping or replicate structure of your experiment. To enter experiment parameters click "Add Parameter...". You can enter as many parameters as you want, but only the first two parameters are used for analysis in this Guided Workflow. Other parameters can be used when analyzing your data with the features available from the Workflow Browser. You can also edit and re-order parameters and parameter values here.
Displaying 6 sample(s) with 1 experiment parameter(s). To change, use the button controls below.
Samples
Treatment
US22502705_251209747382_Untreated.txt Untreated
US22502705_251209747387_Untreated.txt Untreated
US22502705_251209747392_Treated.txt Treated
US22502705_251209747393_Treated.txt Treated
US22502705_251209747394_Untreated.txt Untreated
US22502705_251209747404_Treated.txt Treated
Add Parameter...
Edit Parameter...
Delete Parameter
Help
<< Back
Next >>
Finish
Cancel

Figure 21 *Experiment Grouping of the Agilent Expression Single Color Demo sample data*

3. Review the sample quality in QC on samples in the **Analysis: Biological Significance (Step 3 of 8)** wizard.

This step provides the first view of the data using a Principal Component Analysis (PCA). PCA lets you assess the data by viewing a 3D scatter plot of the calculated principal components. The PCA scores are shown in each of the selection boxes located along the bottom of the 3D PCA Scores window. A higher score indicates that the principal component contains more of the variability of the data. The components generated in the 3D PCA Scores graph are represented in the X, Y, and Z axes and are numbered 1, 2, 3 ... in order of their decreasing significance.

Principal component analysis: The mathematical process by which data containing a number of potentially correlated variables is transformed into a data set in relation to a smaller number of variables called principal components that account for the most variability in the data. The result of the data transformation leads to the identification of the best explanation of the variance in the data, e.g. identification of the components in the data that contain the meaningful information providing differentiation.

Principal component: Transformed data into axes, principal components, so that the patterns between the axes most closely describe the relationships between the data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The principal components are viewed and interpreted in 3D graphical axes with additional dimensions represented by different colors and/or shapes representing the parameter names.

a Review the QC on samples results.

b Click **Next**.

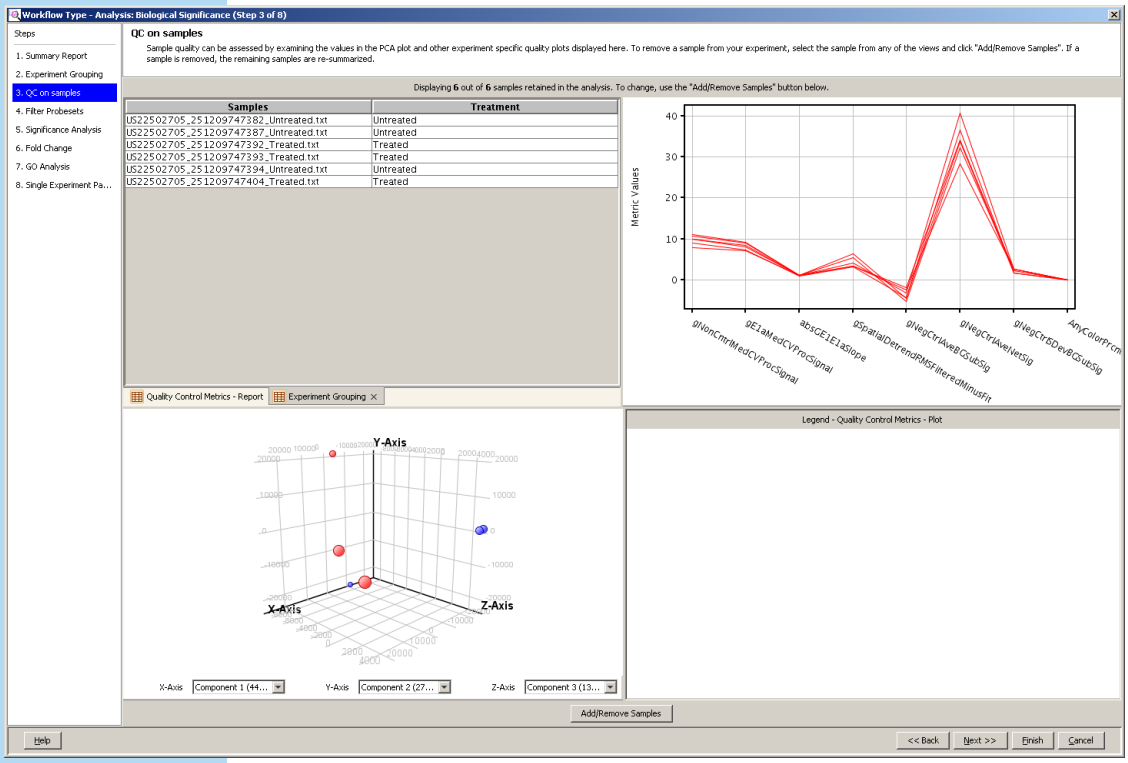

Workflow Type - Analysis: Biological Significance (Step 3 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
QC on samples
Sample quality can be assessed by examining the values in the PCA plot and other experiment specific quality plots displayed here. To remove a sample from your experiment, select the sample from any of the views and click "Add/Remove Samples". If a sample is removed, the remaining samples are re-summarized.
Displaying 6 out of 6 samples retained in the analysis. To change, use the "Add/Remove Samples" button below.
Samples
Treatment
US22502705_251209747382_Untreated.txt Untreated
US22502705_251209747387_Untreated.txt Untreated
US22502705_251209747392_Treated.txt Treated
US22502705_251209747393_Treated.txt Treated
US22502705_251209747394_Untreated.txt Untreated
US22502705_251209747404_Treated.txt Treated
Metric Values
Quality Control Metrics - Report
Experiment Grouping
Legend - Quality Control Metrics - Plot
Y-Axis
X-Axis
Z-Axis
Add/Remove Samples
Help
<< Back
Next >>
Finish
Cancel

Figure 22 *QC on samples for the Agilent Expression Single Color Demo sample data*

4. Review the Filter Probesets in the **Analysis: Biological Significance (Step 4 of 8)** wizard.

a Review the Filter Probesets results.

b Click **Re-run Filter**.

c Mark **Detected** and **Not Detected** as Acceptable Flags.

d Click **OK**.

e Click **Next**.

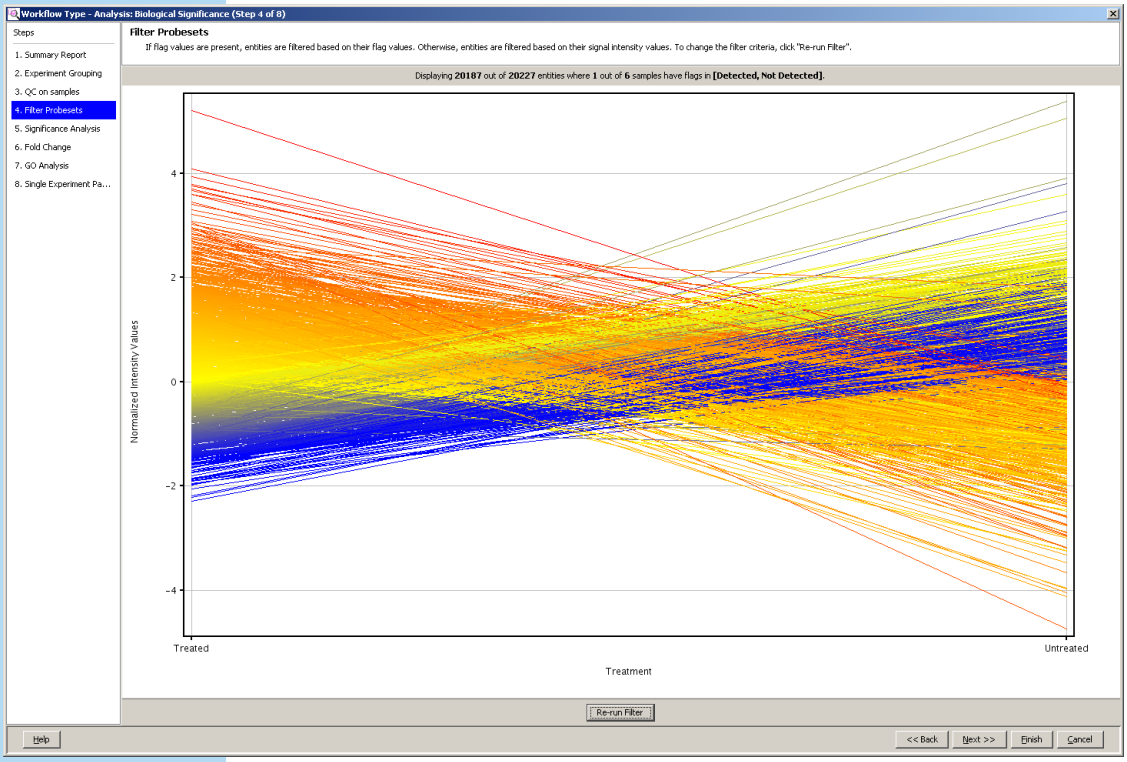


Figure 23 *Filter Probesets for the Agilent Expression Single Color Demo sample data*

5. Review the Significance Analysis in the **Analysis: Biological Significance (Step 5 of 8)** wizard.

a Review the Significance Analysis results.

b Click and move the **Corrected p-value cut-off** slider or type in the p-value cut-off value and press the **Enter** key. The default value is 0.05. The results in the display window are automatically updated.

c Selected **[Treated]** for the **Control Group**.

d Click **Next**.

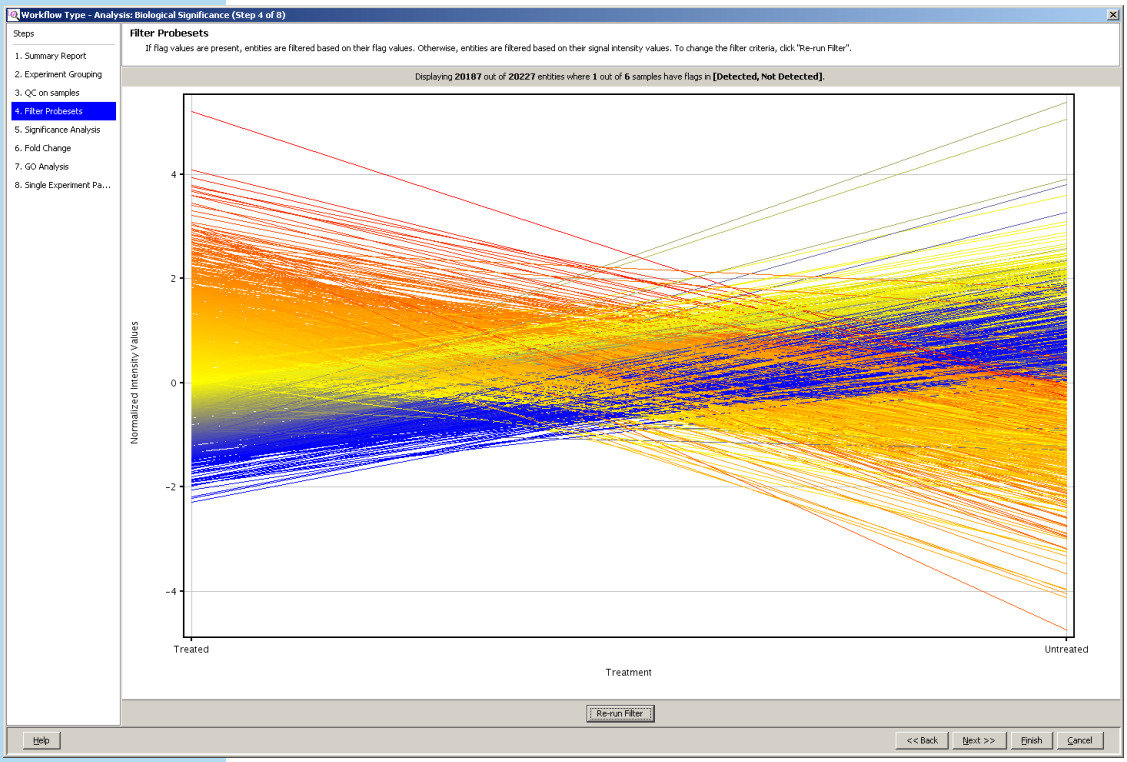


Figure 24 *Significance Analysis for the Agilent Expression Single Color Demo sample data*

6. Review the Fold Change in the **Analysis: Biological Significance (Step 6 of 8)** wizard.

a Review the Fold Change results.

b Click and move the **Fold change p-value cut-off** slider or type in the p-value cut-off value and press the **Enter** key. The default value is 2.0. The results in the display window are automatically updated.

c Click **Next**.

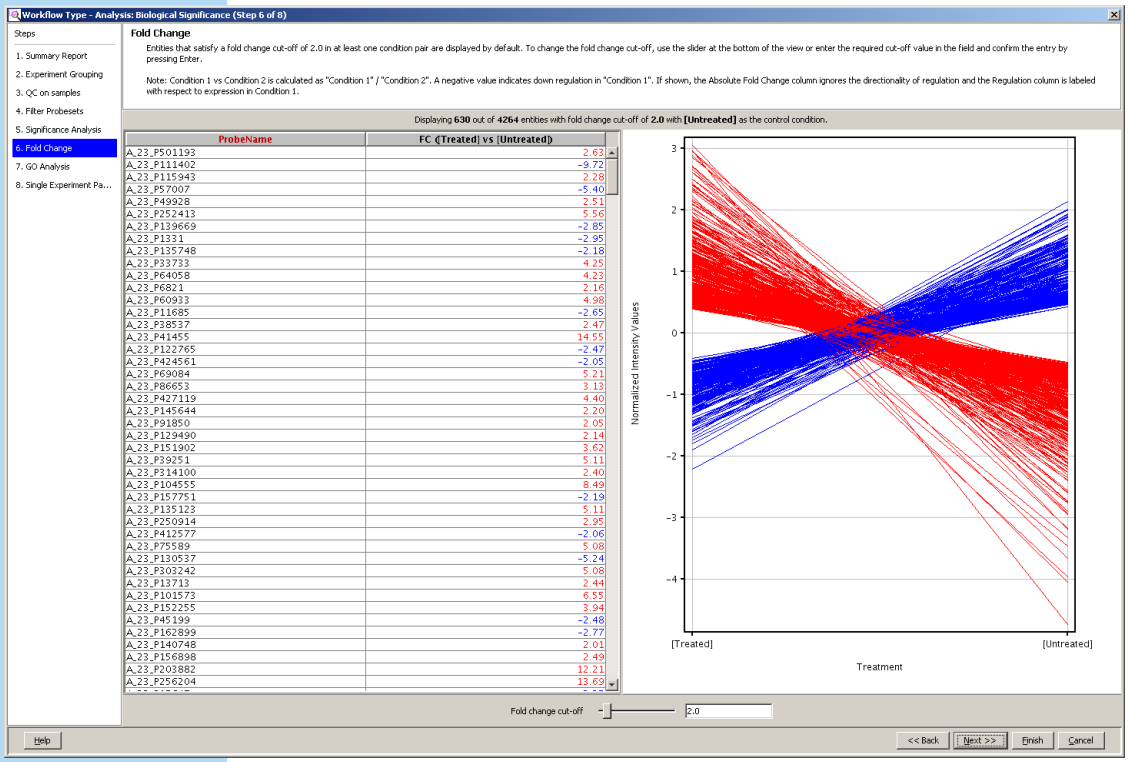


Figure 25 *Fold Change for the Agilent Expression Single Color Demo sample data*

7. Review the GO Analysis in the **Analysis: Biological Significance (Step 7 of 8)** wizard.

a Review the GO Analysis results.

b Click and move the **corrected p-value cut-off** slider or type in the p-value cut-off value and press the **Enter** key. The default value is 0.1. The results in the display window are automatically updated.

c Click **Next**. A progress dialog box is displayed while the Single Experiment Pathway Analysis is performed.

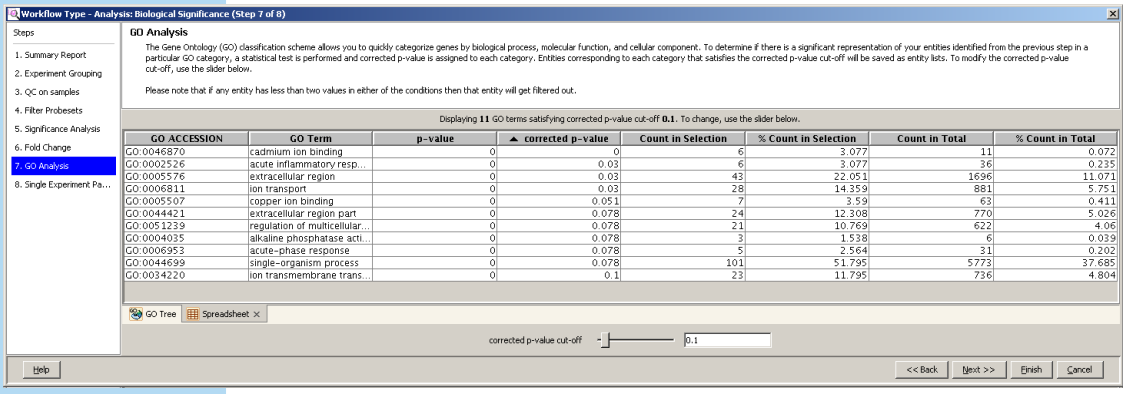

Workflow Type - Analysis: Biological Significance (Step 7 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
GO Analysis
The Gene Ontology (GO) classification scheme allows you to quickly categorize genes by biological process, molecular function, and cellular component. To determine if there is a significant representation of your entities identified from the previous step in a particular GO category, a statistical test is performed and corrected p-value is assigned to each category. Entities corresponding to each category that satisfies the corrected p-value cut-off will be saved as entity lists. To modify the corrected p-value cut-off, use the slider below.
Please note that if any entity has less than two values in either of the conditions then that entity will get filtered out.
Displaying 11 GO terms satisfying corrected p-value cut-off 0.1. To change, use the slider below.
GO ACCESSION
GO Term
p-value
corrected p-value
Count in Selection
% Count in Selection
Count in Total
% Count in Total
GO:0046870 cadmium ion binding 0 0 6 3.077 11 0.072
GO:0002526 acute inflammatory resp... 0 0.03 6 3.077 36 0.235
GO:0005576 extracellular region 0 0.03 43 22.051 1696 11.071
GO:0006811 ion transport 0 0.03 28 14.359 881 5.751
GO:0005507 copper ion binding 0 0.051 7 3.59 63 0.411
GO:0044421 extracellular region part 0 0.078 24 12.308 770 5.026
GO:0051239 regulation of multicellular... 0 0.078 21 10.769 622 4.06
GO:0004035 alkaline phosphatase acti... 0 0.078 3 1.538 6 0.039
GO:0006953 acute-phase response 0 0.078 5 2.564 31 0.202
GO:0044699 single-organism process 0 0.078 101 51.795 5773 37.685
GO:0034220 ion transmembrane trans... 0 0.1 23 11.795 736 4.804
GO Tree
Spreadsheet
corrected p-value cut-off
0.1
Help
<< Back
Next >>
Finish
Cancel

Figure 26 *GO Analysis for the Agilent Expression Single Color Demo sample data*

8. Review the Single Experiment Pathway Analysis in the **Analysis: Biological Significance (Step 8 of 8)** wizard.

a Review the Single Experiment Pathway Analysis results.

b Click **Next**. The pathway list is saved.

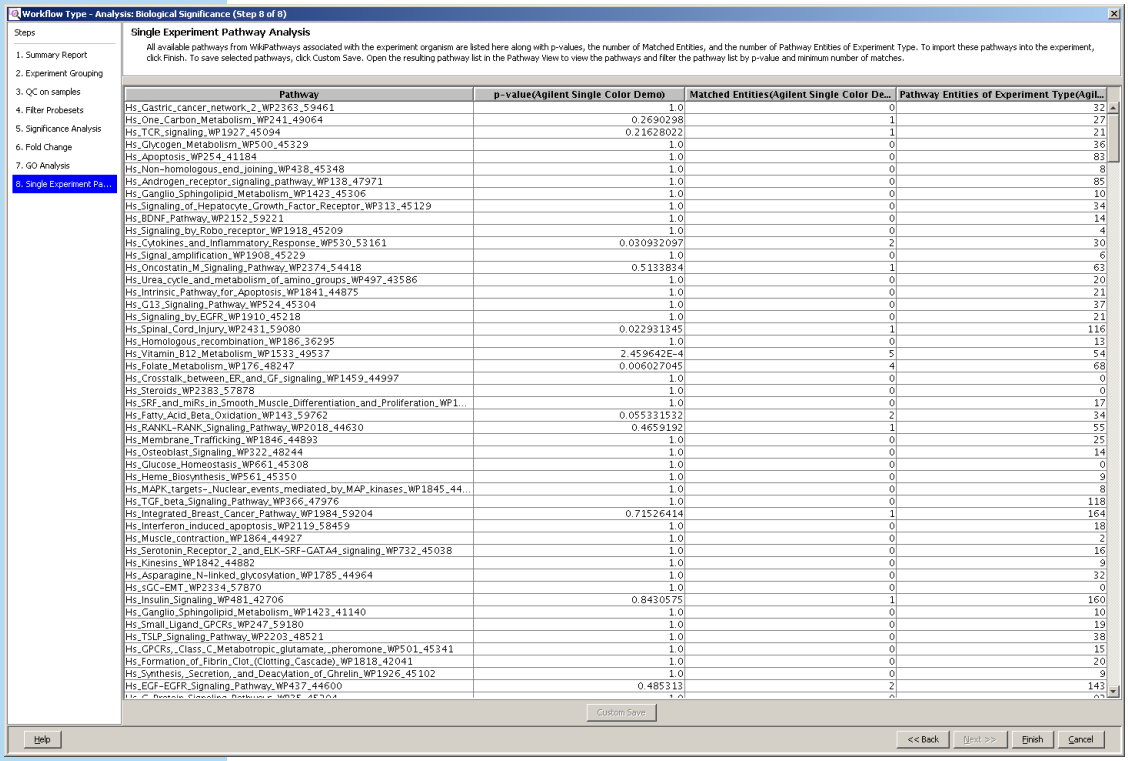

Workflow Type - Analysis: Biological Significance (Step 8 of 8)
Steps
1. Summary Report
2. Experiment Grouping
3. QC on samples
4. Filter Probesets
5. Significance Analysis
6. Fold Change
7. GO Analysis
8. Single Experiment Pa...
Single Experiment Pathway Analysis
All available pathways from WikiPathways associated with the experiment organism are listed here along with p-values, the number of Matched Entities, and the number of Pathway Entities of Experiment Type. To import these pathways into the experiment, click Finish. To save selected pathways, click Custom Save. Open the resulting pathway list in the Pathway View to view the pathways and filter the pathway list by p-value and minimum number of matches.
Pathway
p-value(Agilent Single Color Demo)
Matched Entities(Agilent Single Color De...
Pathway Entities of Experiment Type(Agil...
Hs_Gastric_cancer_network_2_WP2363_59461 1.0 0 32
Hs_One_Carbon_Metabolism_WP241_49064 0.2690298 1 27
Hs_TCR_signaling_WP1927_45094 0.21628022 1 21
Hs_Glycogen_Metabolism_WP500_45329 1.0 0 36
Hs_Apoptosis_WP254_41184 1.0 0 83
Hs_Non-homologous_end_joining_WP438_45348 1.0 0 8
Hs_Androgen_receptor_signaling_pathway_WP138_47971 1.0 0 85
Hs_Ganglio_Sphingolipid_Metabolism_WP1423_45306 1.0 0 10
Hs_Signaling_of_Hepatocyte_Growth_Factor_Receptor_WP313_45129 1.0 0 34
Hs_BDNF_Pathway_WP2152_59221 1.0 0 14
Hs_Signaling_by_Robo_receptor_WP1918_45209 1.0 0 4
Hs_Cytokines_and_Inflammatory_Response_WP530_53161 0.030932097 2 30
Hs_Signal_amplification_WP1908_45229 1.0 0 6
Hs_Oncostatin_M_Signaling_Pathway_WP2374_54418 0.5133834 1 63
Hs_Urea_cycle_and_metabolism_of_amino_groups_WP497_43586 1.0 0 20
Hs_Intrinsic_Pathway_for_Apoptosis_WP1841_44875 1.0 0 21
Hs_G13_Signaling_Pathway_WP524_45304 1.0 0 37
Hs_Signaling_by_EGFR_WP1910_45218 1.0 0 21
Hs_Spinal_Cord_Injury_WP2431_59080 0.022931345 1 116
Hs_Homologous_recombination_WP186_36295 1.0 0 13
Hs_Vitamin_B12_Metabolism_WP1533_49537 2.459642E-4 5 54
Hs_Folate_Metabolism_WP176_48247 0.006027045 4 68
Hs_Crosstalk_between_ER_and_GF_signaling_WP1459_44997 1.0 0 0
Hs_Steroids_WP2383_57878 1.0 0 0
Hs_SRF_and_miRs_in_Smooth_Muscle_Differentiation_and_Proliferation_WP1... 1.0 0 17
Hs_Fatty_Acid_Beta_Oxidation_WP143_59762 0.055331532 2 34
Hs_RANKL-RANK_Signaling_Pathway_WP2018_44630 0.4659192 1 55
Hs_Membrane_Trafficking_WP1846_44893 1.0 0 25
Hs_Osteoblast_Signaling_WP322_48244 1.0 0 14
Hs_Glucose_Homeostasis_WP661_45308 1.0 0 0
Hs_Heme_Biosynthesis_WP561_45350 1.0 0 9
Hs_MAPK_targets-_Nuclear_events_mediated_by_MAP_kinases_WP1845_44... 1.0 0 8
Hs_TGF_beta_Signaling_Pathway_WP366_47976 1.0 0 118
Hs_Integrated_Breast_Cancer_Pathway_WP1984_59204 0.71526414 1 164
Hs_Interferon_induced_apoptosis_WP2119_58459 1.0 0 18
Hs_Muscle_contraction_WP1864_44927 1.0 0 2
Hs_Serotonin_Receptor_2_and_ELK-SRF-GATA4_signaling_WP732_45038 1.0 0 16
Hs_Kinesins_WP1842_44882 1.0 0 9
Hs_Asparagine_N-linked_glycosylation_WP1785_44964 1.0 0 32
Hs_sGC-EMT_WP2334_57870 1.0 0 0
Hs_Insulin_Signaling_WP481_42706 0.8430575 1 160
Hs_Ganglio_Sphingolipid_Metabolism_WP1423_41140 1.0 0 10
Hs_Small_Ligand_GPCRs_WP247_59180 1.0 0 19
Hs_TSLP_Signaling_Pathway_WP2203_48521 1.0 0 38
Hs_GPCRs,_Class_C_Metabotropic_glutamate,_pheromone_WP501_45341 1.0 0 15
Hs_Formation_of_Fibrin_Clot_(Clotting_Cascade)_WP1818_42041 1.0 0 20
Hs_Synthesis,_Secretion,_and_Deacylation_of_Ghrelin_WP1926_45102 1.0 0 9
Hs_EGF-EGFR_Signaling_Pathway_WP437_44600 0.485313 2 143
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 27 *Single Experiment Pathway Analysis for the Agilent Expression Single Color Demo sample data*

You are now in the advanced workflow mode and have access to all features available in Mass Profiler Professional through the Workflow Browser. The imported and analyzed **Agilent Expression Single Color Demo** sample data is displayed in MPP similar to Figure 28 on page 34.

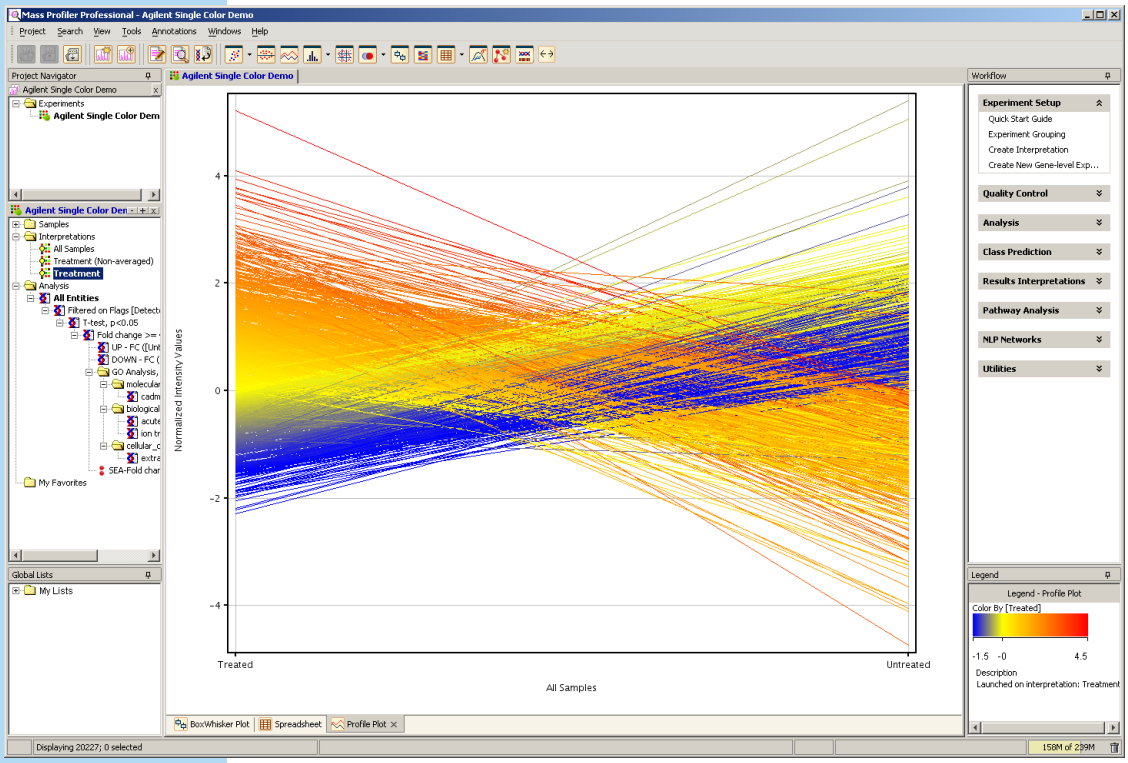
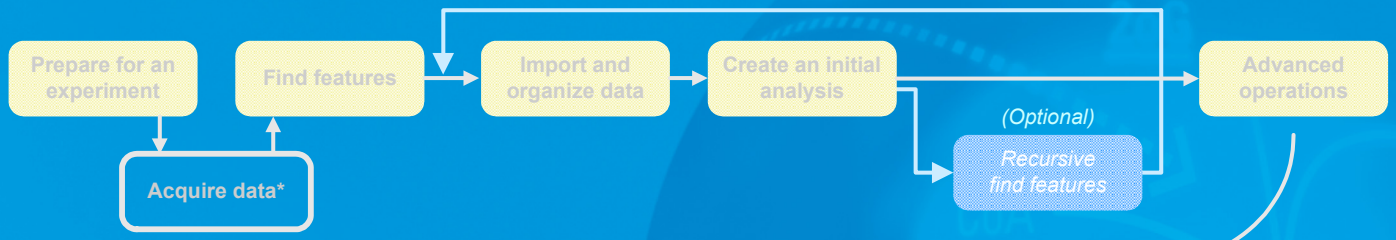
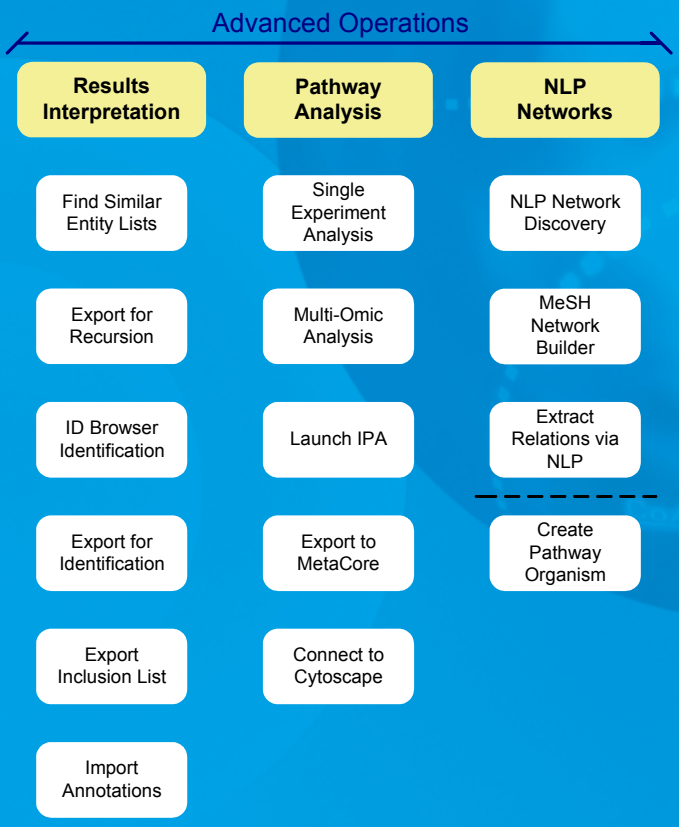


Figure 28 *The Agilent Expression Single Color Demo sample data after importing and performing a biological significance analysis*

Integrated Biology operations

Mass Profiler Professional enables you to analyze data from different high-throughput technologies like genomics, transcriptomics, proteomics, and metabolomics and it also allows you to compare data from these different experiment types in the same project.





* **Acquire data** *is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Overview of operations 36
Results Interpretation 38
Pathway Analysis 61
NLP Networks 92

Agilent Technologies

Overview of operations

Mass Profiler Professional is one of the solutions developed by Agilent to facilitate multi-omic data analysis. The operations available in the Workflow Browser of Mass Profiler Professional provide the tools necessary for analyzing features from your mass spectrometry data depending upon the need and aim of the analysis, the experimental design, and the focus of the study. This helps you create different interpretations to carry out the analysis based on the different filtering, normalization, and standard statistical methods.

Regardless of your personal expertise, the **Analysis: Significance Testing and Fold Change** workflow provides you with quality control to your analysis that improves your results. When you begin your data analysis using Mass Profiler Professional, it is recommended that you follow the procedures in the chapter "Create an initial analysis" in the *Agilent Metabolomics Workflow - Discovery Workflow Guide* before proceeding with the operations available in the integrated biology operations. When you click **Finish** during "Create an initial analysis" (see Figure 5 on page 17) Mass Profiler Professional automatically makes the operations available under the Workflow Browser; you have access to all available operations.

Only some of the operations available in the Workflow Browser are documented in this workflow guide *Integrated Biology with Mass Profiler Professional - Workflow Guide*. This workflow documents the operations that are most relevant to performing your integrated biology analysis:

- Results Interpretations (see "Results Interpretation" on page 38)
- Pathway Analysis (see "Pathway Analysis" on page 61)
- NLP Networks (see "NLP Networks" on page 92)

The *Agilent Metabolomics Workflow - Discovery Workflow Guide* documents the general experimental, data quality, and statistical analysis operations:

- Experiment Setup
- Quality Control
- Analysis

The operations associated with Class Prediction and utilities are documented in *Class Prediction with Mass Profiler Professional - Workflow Guide.*

More information regarding any of the operations available in the Workflow Browser is found in the *Mass Profiler Professional User Manual*.

Layout of the Mass Profiler Professional screen

After you have "imported and organized your data" and then "created an initial differential analysis," MPP places you in the advanced workflow mode where you have access to all features available in Mass Profiler Professional through the Workflow Browser. If you are using the two-variable experiment data set, or similar, you see a display similar to that shown in Figure 29 on page 37. You are ready to perform the integrated biology operations.

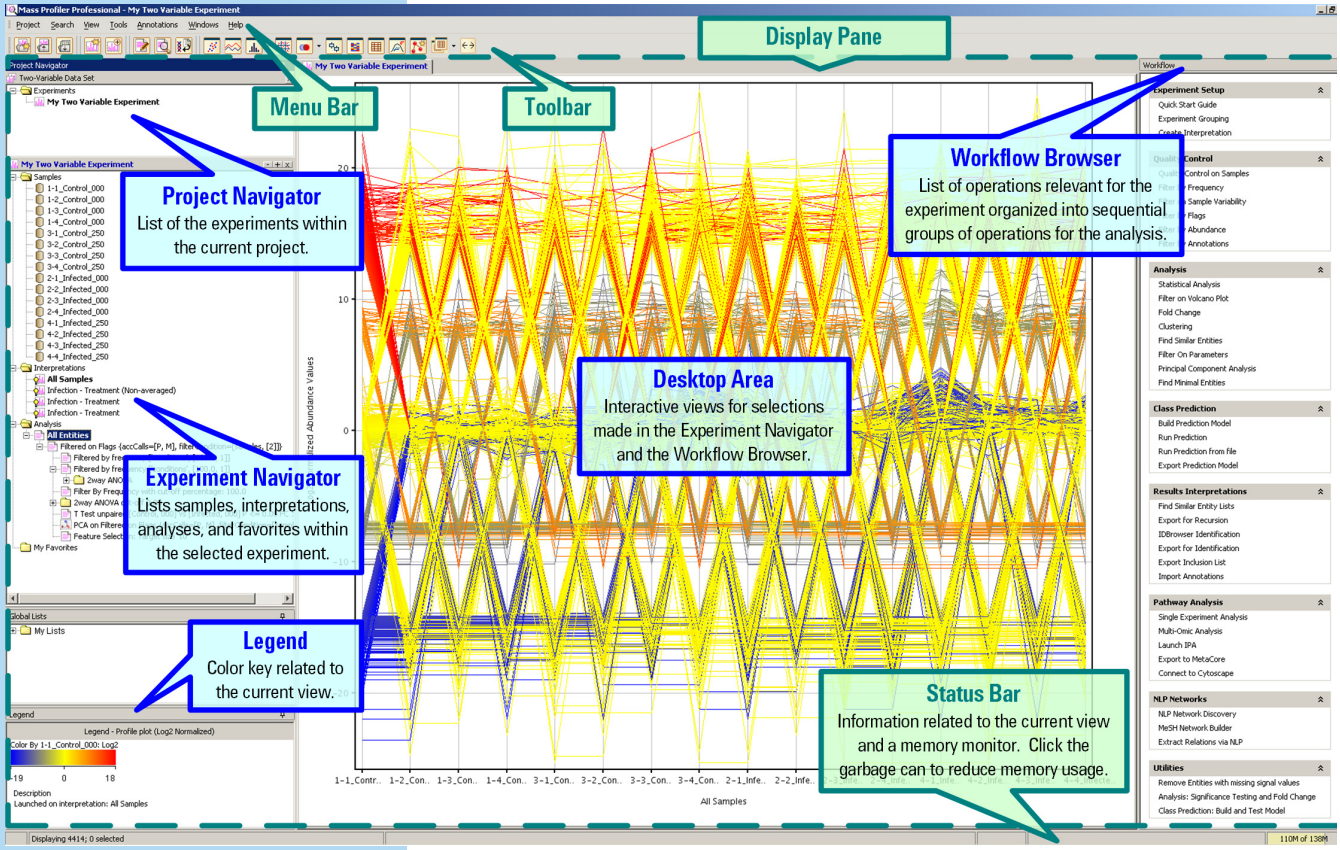
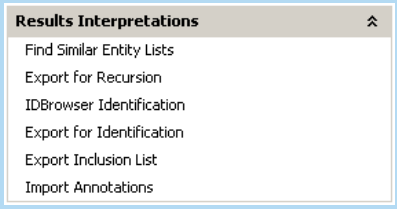

Display Pane
Menu Bar
Toolbar
Project Navigator
List of the experiments within the current project.
Workflow Browser
List of operations relevant for the experiment organized into sequential groups of operations for the analysis.
Desktop Area
Interactive views for selections made in the Experiment Navigator and the Workflow Browser.
Experiment Navigator
Lists samples, interpretations, analyses, and favorites within the selected experiment.
Legend
Color key related to the current view.
Status Bar
Information related to the current view and a memory monitor. Click the garbage can to reduce memory usage.

Figure 29 *The main functional areas of Mass Profiler Professional illustrated using the "Two-variable experiment" data set*

Results Interpretation



With the operations available in Results Interpretation, you can analyze and refine the entities and entity lists that were created during your experimental analyses. Results Interpretation consists of six operations:

- "Find Similar Entity Lists" on page 38
- "Export for Recursion" on page 45
- "ID Browser Identification" on page 47
- "Export for Identification" on page 54
- "Export Inclusion List" on page 55
- "Import Annotations" on page 59

Entity Lists

Entity lists contain the compounds (entities) that meet the conditions specified in each experiment performed on your data. Entity lists are displayed and accessed in the Experiment Navigator. The Experiment Navigator makes it easy for you to view an entity list's relationship among your experiments and select it for reviewing.

Throughout this workflow the entities in an entity list may be individually referred to as a **metabolite**, **compound**, **feature**, **element**, or **entity** during the various operations.

Find Similar Entity Lists

Similar entity lists are the entity lists in your experiment navigator that contain a significant number of entities in common with a specified source entity list. Similarity among entity lists can also be based on filter criteria such as technology, organism, project, and experiment.

The entity lists that meet your filter parameters are compared to the source entity list to determine if any of the target entity lists contain a significant number of entities in common with the source entity list. Significance is adjusted using a p-value cut-off.

p-value cut-off

For any particular test of significance a p-value may be thought of as the probability of rejecting the null hypothesis when it is in fact true. For a p-value of 0.05 approximately one out of every twenty comparisons results in a false positive analysis (rejection of the null hypothesis when in fact it is true). Thus, if your experiment involves performing 100 comparisons with a p-value of 0.05, we expect five of the comparisons to be false positives. A proper statistical treatment therefore controls the false positive rate for the entire comparison set.

A smaller p-value cut-off reduces the rate of obtaining a false positive or false negative result and therefore reduces the number of comparisons that meet your criteria. A larger p-value cut-off increases the rate of obtaining a false positive or false negative result and therefore increases the number of comparisons that meet your criteria.

1. Launch **Find Similar Entity Lists** in the Workflow Browser.

a Click **Find Similar Entity Lists** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Find Similar Entities Lists wizard has three (3) steps plus additional steps involved in choosing your entity list using the EntityList Search Wizard. The steps that you use depending on how you select the target entity lists in the first step of the wizard (see Figure 30). When **Custom** is selected as the **Target entity lists** the **EntityList Search Wizard** is used to input the additional target criteria.

The new entity list is placed in the Analysis folder within the Experiment Navigator. More than one entity list may be created from your analysis.

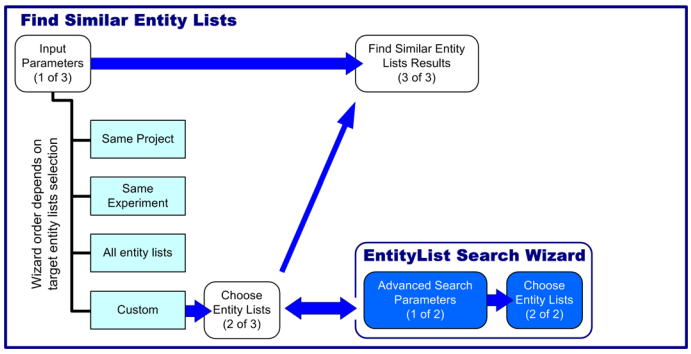

Find Similar Entity Lists
Input Parameters (1 of 3)
Find Similar Entity Lists Results (3 of 3)
Wizard order depends on target entity lists selection
Same Project
Same Experiment
All entity lists
Custom
Choose Entity Lists (2 of 3)
EntityList Search Wizard
Advanced Search Parameters (1 of 2)
Choose Entity Lists (2 of 2)

Figure 30 *Flow chart of the Find Similar Entity Lists wizard.*

2. Select the input parameters in **Find Similar Entity Lists (Step 1 of 3)**.

a Click **Choose** to select the **Entity list** that you want to use as the source for finding similar entity lists. By default, the active entity list is selected.

b Select a filter for **Target entity lists** to compare to the **Entity list**.

The available filter selections for **Target entity lists** are: **Same project**, **Same Experiment**, **All entity list**, and **Custom**.

c Select an additional filter **Type of targets** if available as an option.

You can change the default value from **All Types** to either **Same Technology** or **Same Organism** when your selection for the **Target entity lists** is **Same Project** or **All entity lists**.

d Click **Next**.

If you select **Custom** for the **Target entity lists** (see Figure 32 on page 40) proceed to step 3 “Begin entity list search from Find Similar Entity Lists (Step 2 of 3).” on page 40, otherwise proceed to step 7 “Select and save entity lists based on significance in Find Similar Entity Lists (Step 3 of 3).” on page 43.

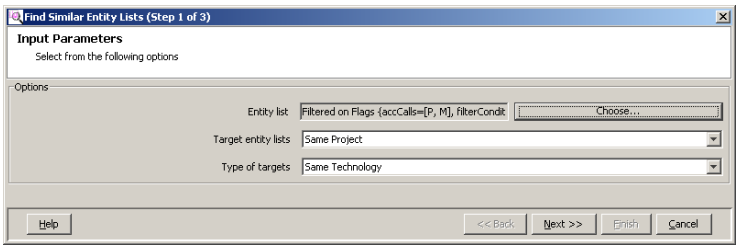


Figure 31 *Input Parameters page (Find Similar Entity Lists (Step 1 of 3))*

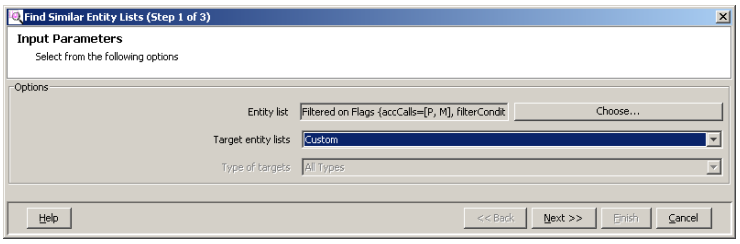


Figure 32 *Input Parameters page with* ***Custom*** *selected for the* ***Target entity lists*** *(Find Similar Entity Lists (Step 1 of 3))*

3. Begin entity list search from **Find Similar Entity Lists (Step 2 of 3)**.

a Click **Choose EntityList(s)** to begin the **EntityList Search Wizard**.

This step is only performed if you select **Custom** for the **Target entity lists** (see Figure 32). The entity list table is empty until you complete the **EntityList Search Wizard** in the following steps.

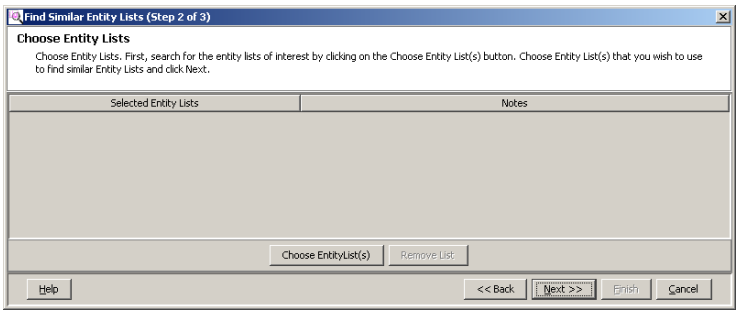

Find Similar Entity Lists (Step 2 of 3)
Choose Entity Lists
Choose Entity Lists. First, search for the entity lists of interest by clicking on the Choose Entity List(s) button. Choose Entity List(s) that you wish to use to find similar Entity Lists and click Next.
Selected Entity Lists
Notes
Choose EntityList(s)
Remove List
Help
<< Back
Next >>
Finish
Cancel

Figure 33 *Input Parameters page (Find Similar Entity Lists (Step 1 of 3))*

4. Enter entity list filter criteria in **EntityList Search Wizard (Step 1 of 2)**.

Build the entity list filter criteria, referred to as a search query, to find the entity lists to compare to the source entity list specified as the **Entity List** in step 2 on page 39.

Since the conditions and search values you enter depends on the selected search field, Table 1 provides you with an overview of the available parameters to build your entity list search query.

Table 1 *List of available parameters to build your search query.*

Search Field	Applicable Condition Group	Condition	Search Value
		Group A	
Creation Date	A	=	Enter value or select date
Global List	C	≥	Enter value or select date
Mark	B	≤	Enter value or select date
Modified Date	A		
Name	B	***Group B***	
Notes	B	equals	Enter text
Number of Entities	A	starts with	Enter text
System Object	B	ends with	Enter text
Technology	B	includes	Enter text
Type	B		
		Group C	
		equals	Select true

a Select a **Search Field**.

b Select a **Condition**. The available conditions depend on the **Search Field** selection as shown in Table 1.

c Enter a **Search Value**, or select a date or condition depending on your **Search Field** selection as shown in Table 1.

d Select the **AND** or **OR** operator in **Combine search conditions by** if your entity list filter includes criteria for more than one **Search Field**.

If your criteria has only a single **Search Field**, or if this is your last combined **Search Field** row, proceed to step f.

e Repeat step a through step c for each of your combined filter criteria.

f Enter a value for **Max results per page** to adjust how you plan to review the entity lists on the next step of the wizard.

g Click **Next**.

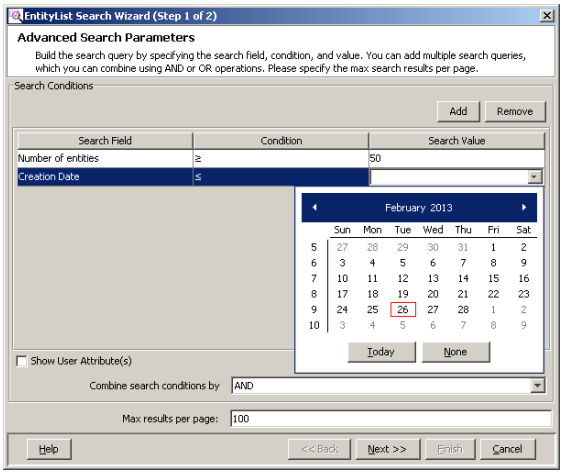


Figure 34 *Advanced Search Parameters page (EntityList Search Wizard (Step 1 of 2)) with two filter criteria, the last one requiring the selection of a date*

---

5. Review the search results in **EntityList Search Wizard (Step 2 of 2)**.

a Review the entity lists that met your search criteria.

b Click **Back** if you want to adjust and rerun your search criteria.

c Click the **forward** and **back** buttons as necessary to review all of the search results.

d Select any or all of the entity lists to return the entity list(s) to the page **Find Similar Entity Lists (Step 2 of 2)**. When an entity list is selected the row is highlighted.

Select a continuous range of entity lists - click on the first file and press **Shift** and click on the last entity list that includes the range of entity lists you want to select.

Select discontinuous or individual entity list - press **Ctrl** and click on additional entity lists.

Note: If your entity list search results span more than one page and you want to make range and/or individual entity list selections across multiple pages, click

Back and increase the value for **Max results per page** so that all of the results are on a single page.

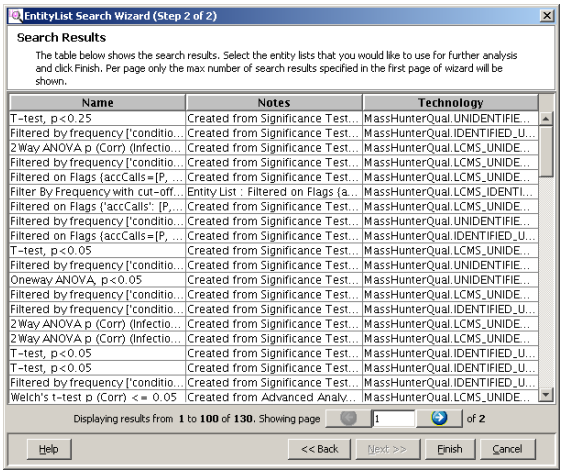

EntityList Search Wizard (Step 2 of 2)
Search Results
The table below shows the search results. Select the entity lists that you would like to use for further analysis and click Finish. Per page only the max number of search results specified in the first page of wizard will be shown.
Name
Notes
Technology
Displaying results from 1 to 100 of 130. Showing page
of 2
Help
<< Back
Next >>
Finish
Cancel

Figure 35 *Search Results page (EntityList Search Wizard (Step 2 of 2))*

e Click **Finish**.

---

6. Choose entity lists in **Find Similar Entity Lists (Step 2 of 3)**.

a Click **Choose EntityList(s)** to rerun the **EntityList Search Wizard** to add additional entity lists.

This step is only performed if you select **Custom** for the **Target entity lists** (see Figure 32). The entity list table is now filled with the entity lists that met your search criteria from the **EntityList Search Wizard**.

b Select one or more entity lists to remove them from further analysis. When an entity list is selected the row is highlighted. See "Review the search results in EntityList Search Wizard (Step 2 of 2)." on page 41 for selecting multiple rows.

c Click **Remove List** to remove the selected entity lists from further analysis.

d Click **Next**.

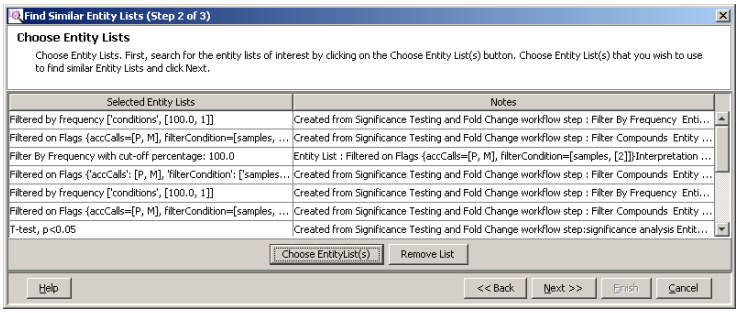

Find Similar Entity Lists (Step 2 of 3)
Choose Entity Lists
Choose Entity Lists. First, search for the entity lists of interest by clicking on the Choose Entity List(s) button. Choose Entity List(s) that you wish to use to find similar Entity Lists and click Next.
Selected Entity Lists
Notes
Choose EntityList(s)
Remove List
Help
<< Back
Next >>
Finish
Cancel

Figure 36 *Choose Entity Lists page (Find Similar Entity Lists (Step 2 of 3))*

7. Select and save entity lists based on significance in **Find Similar Entity Lists (Step 3 of 3)**.

a Review your results.

b *(Optional)* Select one or more entity lists to save them as a custom entity list.

Save a custom entity list

1. Click one or more entity lists. See "Review the search results in EntityList Search Wizard (Step 2 of 2)." on page 41 for selecting multiple rows.
2. Click **Custom Save**. This option is only available if one or more entity lists are selected; a selected entity list row is highlighted.

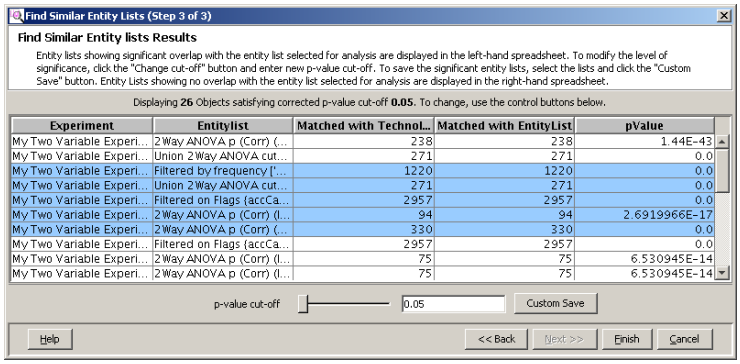

Find Similar Entity Lists (Step 3 of 3)
Find Similar Entity lists Results
Entity lists showing significant overlap with the entity list selected for analysis are displayed in the left-hand spreadsheet. To modify the level of significance, click the "Change cut-off" button and enter new p-value cut-off. To save the significant entity lists, select the lists and click the "Custom Save" button. Entity Lists showing no overlap with the entity list selected for analysis are displayed in the right-hand spreadsheet.
Displaying 26 Objects satisfying corrected p-value cut-off 0.05. To change, use the control buttons below.
Experiment
Entitylist
Matched with Technol...
Matched with EntityList
pValue
My Two Variable Experi... 2Way ANOVA p (Corr) (... 238 238 1.44E-43
My Two Variable Experi... Union 2Way ANOVA cut... 271 271 0.0
My Two Variable Experi... Filtered by frequency ('... 1220 1220 0.0
My Two Variable Experi... Union 2Way ANOVA cut... 271 271 0.0
My Two Variable Experi... Filtered on Flags (accCa... 2957 2957 0.0
My Two Variable Experi... 2Way ANOVA p (Corr) (l... 94 94 2.6919966E-17
My Two Variable Experi... 2Way ANOVA p (Corr) (... 330 330 0.0
My Two Variable Experi... Filtered on Flags (accCa... 2957 2957 0.0
My Two Variable Experi... 2Way ANOVA p (Corr) (l... 75 75 6.530945E-14
My Two Variable Experi... 2Way ANOVA p (Corr) (l... 75 75 6.530945E-14
p-value cut-off
0.05
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 37 *Find Similar Entity Lists Results page (Find Similar Entity Lists (Step 3 of 3))*

3. Add or edit descriptive information that is stored with the saved entity list in the **Name**, **Notes**, and **Experiments** fields on the **Significant EntityLists** page (Figure 39 on page 44).
4. Click **Configure Columns** to add/remove and reorder the columns in the tabular presentation of the entities. This opens the **Select Annotation Columns** dialog box.

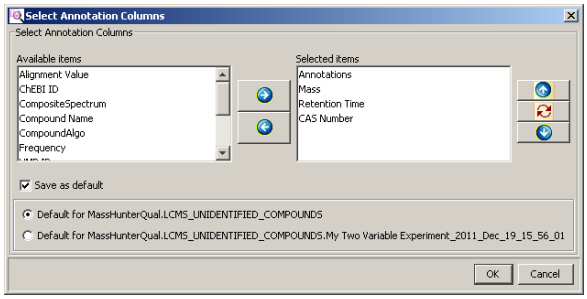


Figure 38 *Select Annotation Columns dialog box*

5. Select column items to add or to remove from the saved entity list.
6. Reorder the selected columns to your preference.
7. Mark **Save as Default** if you would like this configuration to be saved as the default for future save entity list steps.
8. Select the experiment type for your configuration to be applied.
9. Click **OK**.
10. Click **OK**. The entity lists are saved in a folder named "Custom saved Similar Lists" under the source Entity List in the Experiment Navigator.

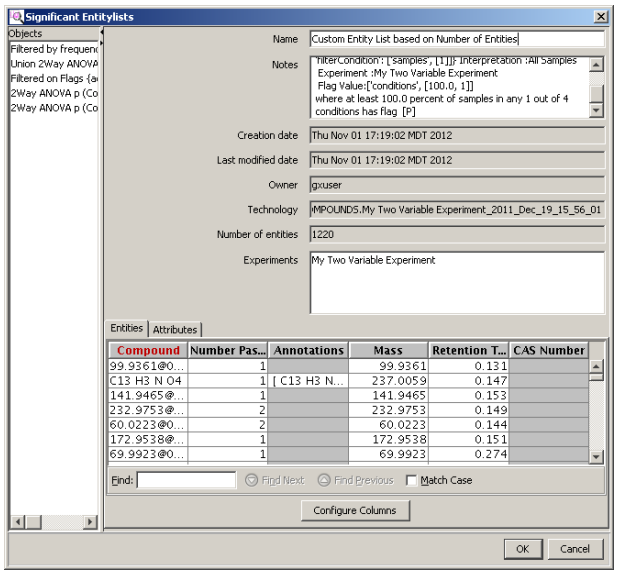


Figure 39 *Saving custom, significant entity lists.*

(End of the optional procedure to select one or more entity lists to save them as a custom entity list)

c Move the slider or type in the **p-value cut-off** value. The default value is 0.05.

Move the slider **p-value cut-off** until the results displayed are satisfactory. Rerun the p-value adjustment several times to develop an understanding of how the p-value cut-off affects your results. A larger p-value passes a larger number of entity lists.

d Click **Back**, make changes to prior parameters, and click **Next** to return to the results until you are satisfied with your analysis.

e Click **Finish**. All of the entity lists shown on the page, whether they are or are not highlighted, are saved in a folder named "Similar Lists satisfying..." under the source Entity List in the Experiment Navigator.

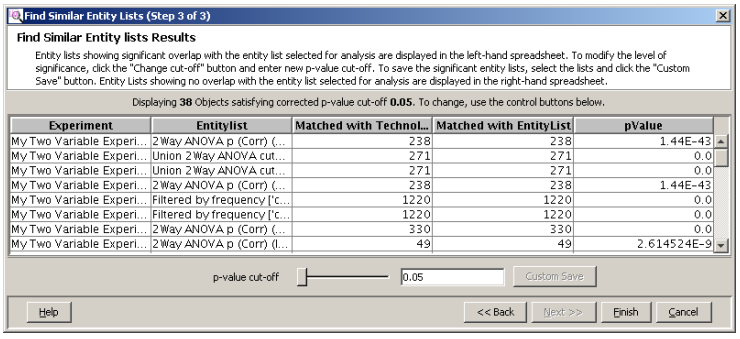

Find Similar Entity Lists (Step 3 of 3)
Find Similar Entity lists Results
Entity lists showing significant overlap with the entity list selected for analysis are displayed in the left-hand spreadsheet. To modify the level of significance, click the "Change cut-off" button and enter new p-value cut-off. To save the significant entity lists, select the lists and click the "Custom Save" button. Entity Lists showing no overlap with the entity list selected for analysis are displayed in the right-hand spreadsheet.
Displaying 38 Objects satisfying corrected p-value cut-off 0.05. To change, use the control buttons below.
Experiment
Entitylist
Matched with Technol...
Matched with EntityList
pValue
My Two Variable Experi... 2Way ANOVA p (Corr) (... 238 238 1.44E-43
My Two Variable Experi... Union 2Way ANOVA cut... 271 271 0.0
My Two Variable Experi... Union 2Way ANOVA cut... 271 271 0.0
My Two Variable Experi... 2Way ANOVA p (Corr) (... 238 238 1.44E-43
My Two Variable Experi... Filtered by frequency ['c... 1220 1220 0.0
My Two Variable Experi... Filtered by frequency ['c... 1220 1220 0.0
My Two Variable Experi... 2Way ANOVA p (Corr) (... 330 330 0.0
My Two Variable Experi... 2Way ANOVA p (Corr) (l... 49 49 2.614524E-9
p-value cut-off
0.05
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 40 *Choose Entity Lists page (Find Similar Entity Lists (Step 3 of 3))*

Export for Recursion

Export the entities in a selected entity list to a CEF file (Compound Exchange Format). The entities exported to a CEF file are used by Agilent MassHunter Qualitative Analysis to find targeted features, the exported entities, from your original sample data files. Recursive feature finding combined with replicate samples improves the statistical accuracy of your analysis and reduces the potential for obtaining a false positive or false negative answer to your hypothesis.

Recursive finding

MassHunter Qualitative Analysis Find Compounds by Formula (FbF) typically uses molecular formula information to calculate the ions and isotope patterns derived from the formula as the basis to find features in the sample data file. When the input molecular features consist of mass and retention time, instead of molecular formula, FbF calculates reasonable isotope patterns and uses these patterns with retention time tolerances to find the target features in the sample data files. When the input molecular features are filtered from a find process that was previously untargeted, the molecular features found using this repeated process of finding molecular features is referred to as recursive finding.

Recursive finding consists of three steps:

1. Untargeted **Find Compounds by Molecular Feature** in MassHunter Qualitative Analysis to find your initial entities.
2. Filtering by **Significance Testing and Fold Change** using abundance, retention time, sample variability, flags, frequency, and statistical significance in Mass Profiler Professional to find your most significant entities.
3. Targeted **Find Compounds by Formula** in MassHunter Qualitative Analysis to improve the reliability of finding your features and subsequently improve your statistical analysis accuracy.

1. Launch **Export for Recursion** in the Workflow Browser.

a Click **Export for Recursion** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Export for Recursion operation has one (1) step as shown in Figure 41.

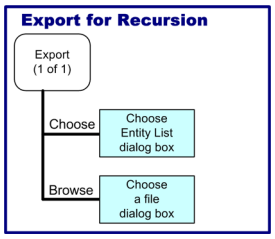


Figure 41 *Flow chart of the Export for Recursion operation.*

2. Select the entity list to export.

a Click **Choose** in the **Export** dialog box.

b Select the entity list to export.

Click an entity list that is at least **Filtered on Flags** from the entity lists in the **Choose Entity List** dialog box. More significance in your analysis is obtained by

selecting an entity list that has at least been filtered by flags to remove "one-hit wonders."

A "one-hit wonder" is a compound that appears in only one sample and is absent from the replicate samples. Therefore, a "one-hit wonder" compound does not provide any utility for statistical analysis and you want to filter such compounds from your analysis.

c Click **OK**.

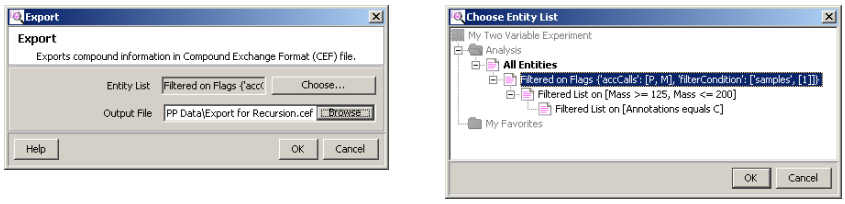


Figure 42 *Export and Choose Entity List dialog boxes*

3. Enter the export file name and folder.

a Click **Browse** in the **Export** dialog box.

Do not type a file name at this location.

b Select the folder or create a new folder for your CEF file in the **Choose a file** dialog box.

c Type the **File name**. For example, you can type `Export for Recursion.cef`.

d Click **Save**.

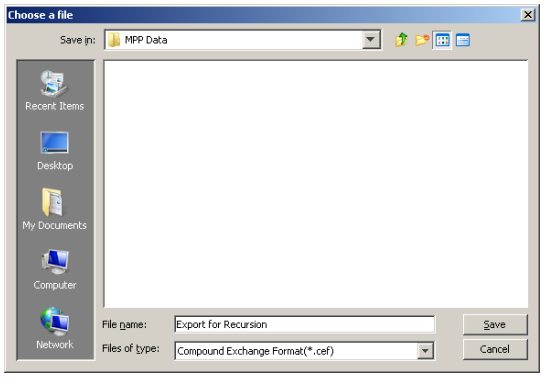


Figure 43 *Choose a file dialog box*

e Click **OK**.

ID Browser Identification

ID Browser identifies and annotates the entities in your selected entity list using LC/MS compound databases (METLIN, pesticides, forensics), GC/MS libraries (NIST and Agilent Fiehn Metabolomics), and empirical formula calculations using Agilent's molecular formula generator (MFG).

When entity identification is completed ID Browser saves and returns and an identified CEF file to Mass Profiler Professional. This CEF is imported into the Mass Profiler Professional experiment and annotations in the selected entity list are updated.

1. Launch **IDBrowser Identification** in the Workflow Browser.

a Click **IDBrowser Identification** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The ID Browser Identification operation has one (1) step within Mass Profiler Professional and three (3) additional steps within ID Browser as shown in Figure 44.

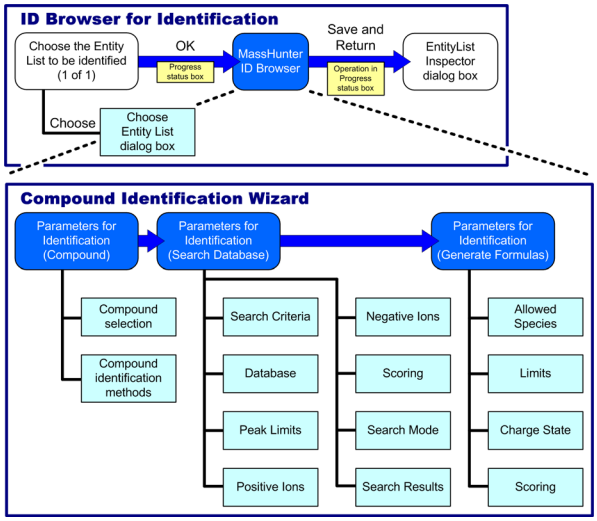


Figure 44 *Flow chart of the ID Browser Identification operation.*

Your entities are initially unidentified as shown in Figure 45 on page 48. When you complete **IDBrowser Identification** your entities appear as shown in Figure 55 on page 54.

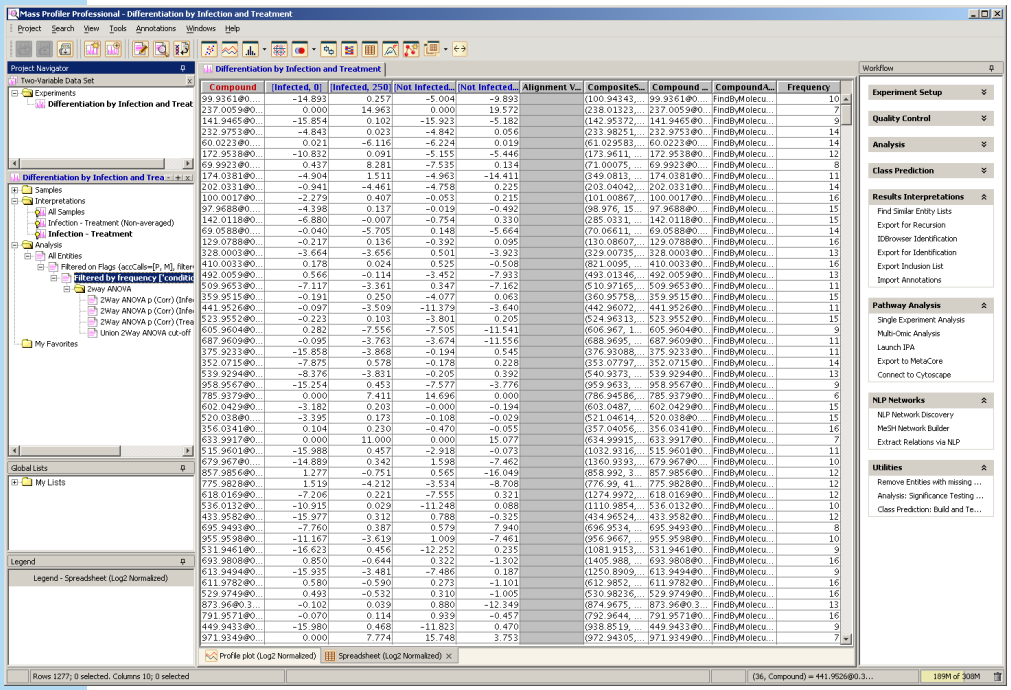


Figure 45 *Spreadsheet view of unidentified entities in MPP before ID Browser.*

2. Select the entity list to identify.

a Click **Choose** in the **Choose the Entity List to be Identified** dialog box.

b Select the entity list to identify.

Since this is an identification operation, you do not need to select an entity list that is at least **Filtered on Flags** from the entity lists in the **Choose Entity List** dialog box.

c Click **OK**.

d Click **OK** to launch ID Browser and transfer the entity list for identification. This action can take extra time and displays a progress status box while ID Browser is starting.

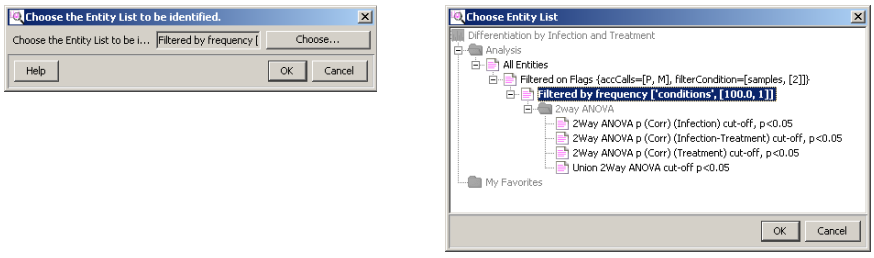


Figure 46 *Choose the Entity List to be Identified and Choose Entity List dialog boxes*

3. Enter the compound selection and identification methods in **Compound Identification Wizard**.

When Mass Profiler Professional launches ID Browser the **Compound Identification Wizard** is automatically started to help you identify your entities. This is the first of two dialog boxes related to this wizard.

a Select **Identify all compounds** for **Compound selection**.

b Mark **Database search** for **Compound identification methods**.

c Mark **Molecular Formula Generator (MFG).**

d Select **Generate formulas only for unidentified compounds**. Generate formulas for the compounds that are not identified by the database search, or the spectral library search, if marked.

e Click **Next**.

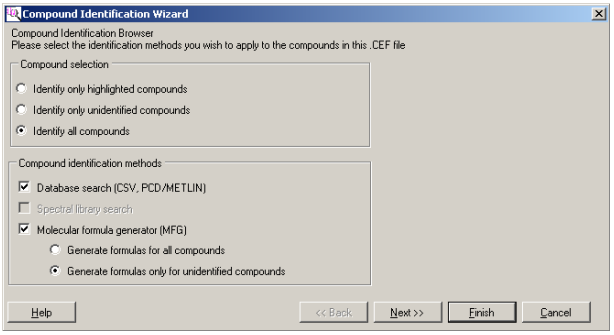


Figure 47 *Parameters related to compound selection and compound identification methods in the Compound Identification Wizard.*

4. Set up the identification techniques in **Compound Identification Wizard**.

The parameters that control the compound identification technique are entered in this second dialog box of the **Compound Identification Wizard**.

a Select **Search Database** under **Identify Compounds**.

b Enter parameters for the **Search Criteria**, **Database**, **Peak Limits**, **Positive Ions**, **Negative Ions**, **Scoring**, **Search Mode**, and **Search Results** tabs similar to that shown in Figure 48, Figure 49 on page 50, and Table 2 on page 50.

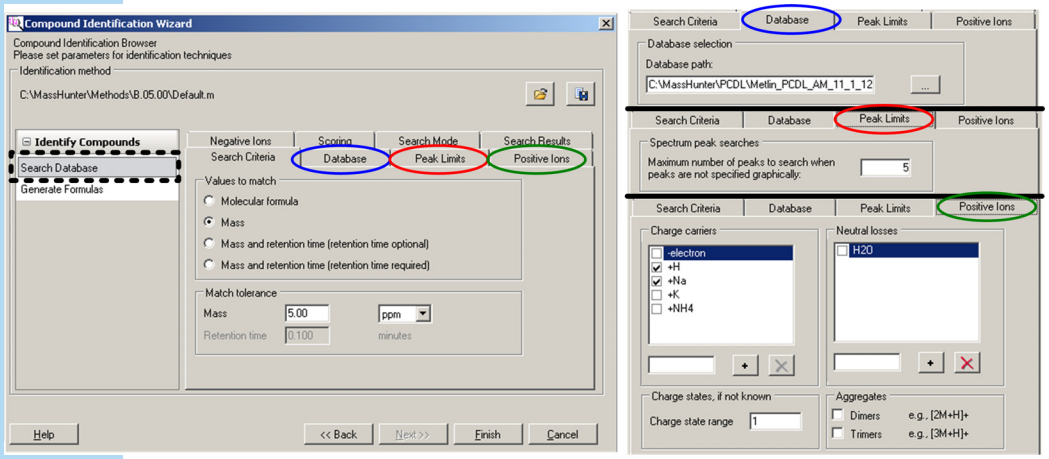


Figure 48 *Parameters for Search Database in the Compound Identification Wizard.*

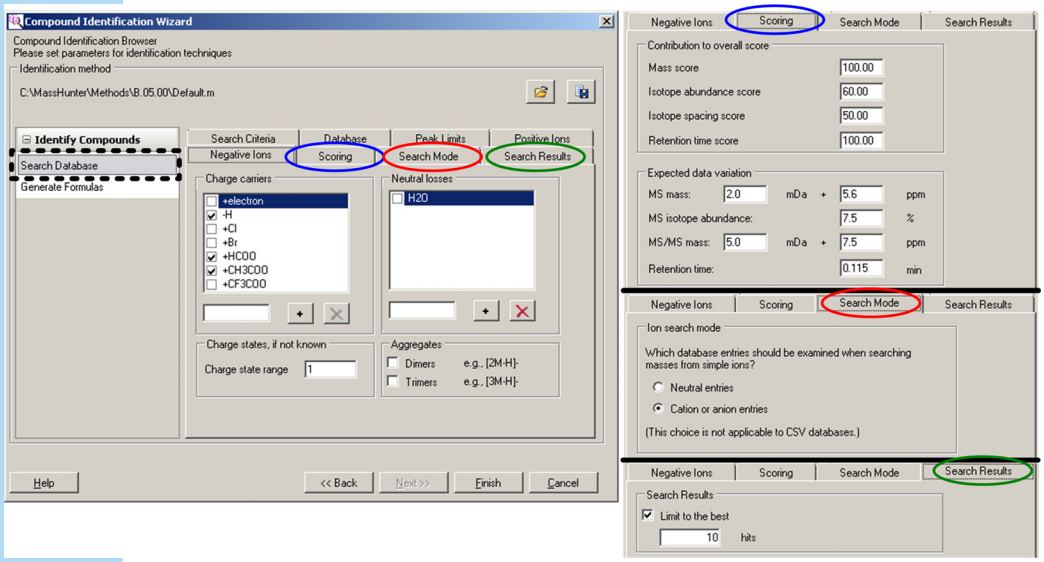


Figure 49 *Parameters for Search Database in the Compound Identification Wizard.*

c Select **Generate Formulas** under **Identify Compounds**.

d Enter parameters for the **Allowed Species**, **Limits**, **Charge State**, and **Scoring** tabs similar to that shown in Figure 50 on page 51 and Table 3 on page 51.

Table 2 ***Search Database*** *Parameters in the Compound Identification Wizard.*

Search Database Parameters		
Tab	Parameter Set	Action
Search Criteria	**Values to match**	Select **Mass**
	Match tolerance	Type **5** and select **ppm** for **Mass**
Database	**Database selection**	Browse and select the MS database
Peak Limits	**Spectrum peak searches**	Type **5** for **Maximum number of peaks to search when peaks are not specified graphically**
Positive Ions	**Charge carriers**	Mark **+H** and **+Na**
	Neutral losses	Clear **H2O**
	Charge states, if not known	Type **1** for **Charge state range**
	Aggregates	Clear **Dimers** and **Trimers**
Negative Ions	**Charge carriers**	Mark **-H**, **+HCOO**, and **+CH3COO**
	Neutral losses	Clear **H2O**
	Charge states, if not known	Type **1** for **Charge state range**
	Aggregates	Clear **Dimers** and **Trimers**
Scoring	**Contribution to overall score**	Type **100** for **Mass score** Type **60** for **Isotope abundance score** Type **50** for **Isotope spacing score** Type **100** for **Retention time score**
	Expected data variation	Type **2.0** mDa and **5.6** ppm for **MS mass** Type **7.5** % for **MS isotope abundance** Type **5.0** mDa and **7.5** ppm for **MS/MS mass** Type **0.115** min for **Retention time**
Search Mode	**Ion search mode**	Select **Cation or anion entries**
Search Results	**Search Results**	Mark **Limit to the best**

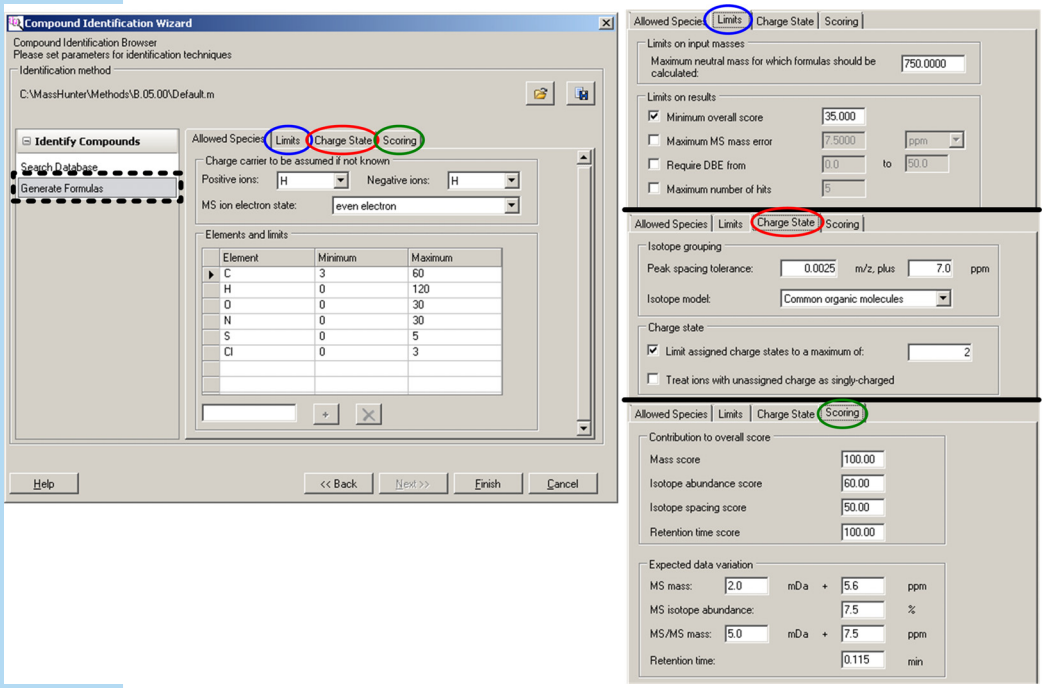


Figure 50 *Parameters for Generate Formula in the Compound Identification Wizard.*

Table 3 ***Generate Formula*** *Parameters in the Compound Identification Wizard.*

Generate Formulas Parameters		
Tab	Parameter Set	Action
Allowed Species	**Charge carriers to be assumed if not known**	Select **H** for **Positive ions**
		Select **H** for **Negative ions**
		Select **even electron** for **MS ion electron state**
	Elements and limits	Type **C 3 60**
		Type **H 0 120**
		Type **O 0 30**
		Type **N 0 30**
		Type **S 0 5**
		Type **Cl 0 3**
Limits	**Limits on input masses**	Type **750** for **Maximum neutral mass for which formulas should be calculated**
	Limits on results	Mark **Minimum overall score** and type **35**
Charge State	**Isotope grouping**	Type **0.0025** m/z and **7.0** ppm for **Peak spacing tolerance**
		Select **Common organic molecules** for **Isotope model**
Scoring	**Contribution to overall score**	Type **100** for **Mass score**
		Type **60** for **Isotope abundance score**
		Type **50** for **Isotope spacing score**
		Type **100** for **Retention time score**
	Expected data variation	Type **2.0** mDa and **5.6** ppm for **MS mass**
		Type **7.5** % for **MS isotope abundance**
		Type **5.0** mDa and **7.5** ppm for **MS/MS mass**
		Type **0.115** min for **Retention time**

e Click **Finish** when you have the method set up for your experiment. ID Browser automatically begins identifying your entities and shows a progress bar.

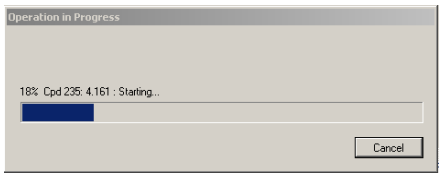


Figure 51 *Progress indication while ID Browser is identifying your entities.*

5. Review your ID Browser results.

When the identification is complete, use the ID Browser interface review your results and make adjustments before returning the identification results to Mass Profiler Professional.

a Review and make adjustments to the entity identifications as necessary.

The ID Browser interface is shown in Figure 52 on page 52. Additional information regarding the use of ID Browser is obtained using **Help** found on the menu bar.

b Click **Save and Return** to export your identified entity list back to your experiment in Mass Profiler Professional.

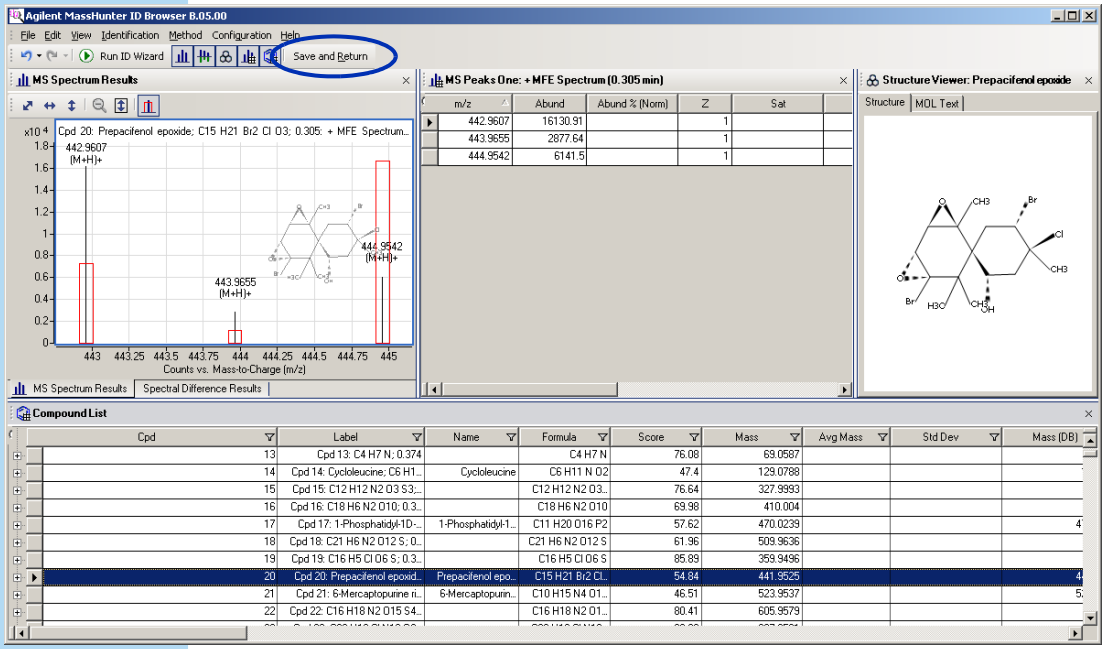


Figure 52 *ID Browser user interface after completing the Compound Identification Wizard.*

6. Review results and enter information in the **Entity-List Inspector** dialog box.

a Review the content and parameters in the **EntityList Inspector** dialog box.

The information and content in the **EntityList Inspector** dialog box are the same for many operations within Mass Profiler Professional that end with a **Save Entity List** page. The figures and description presented in this step are identical to those in other operations. You are referred back to this section when you are prompted

to save your entity list at the completion of other operations available in the Workflow Browser.

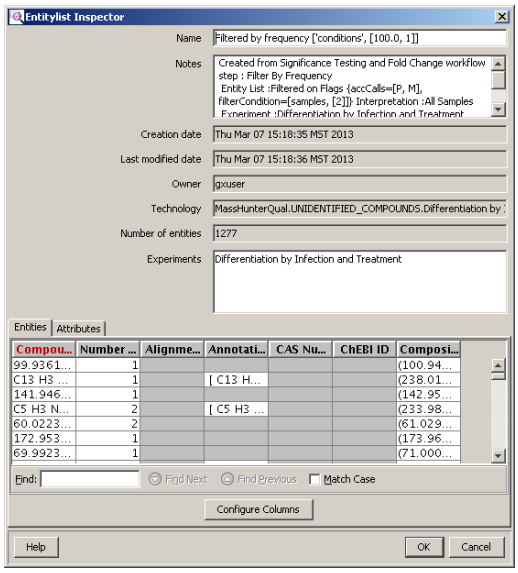


Figure 53 *EntityList Inspector dialog box*

b Add or edit descriptive information that is stored with the saved entity list in the **Name**, **Notes**, and **Experiments** fields (see Figure 53 on page 53).

c Click **Configure Columns** to add/remove and reorder the columns in the tabular presentation of the entities. This opens the **Select Annotation Columns** dialog box (see Figure 54).

d Select column items to add or to remove from the saved entity list.

e Reorder the selected columns to your preference.

f Mark **Save as Default** if you would like this configuration to be saved as the default for future save entity list steps.

g Select the experiment type for your configuration to be applied.

h Click **OK** to exit the **Select Annotation Columns** dialog box.

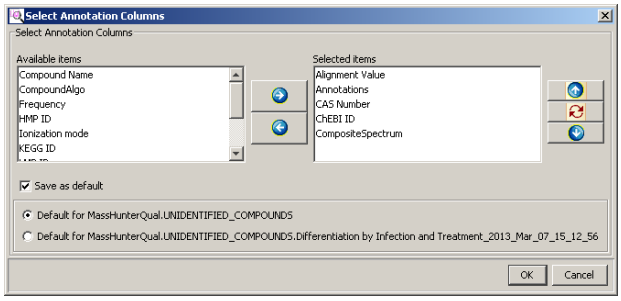


Figure 54 *Select Annotation Columns dialog box*

i Click **OK** to complete the **IDBrowser Identification** operation. At this time your entities in Mass Profiler Professional are identified as shown in Figure 55.

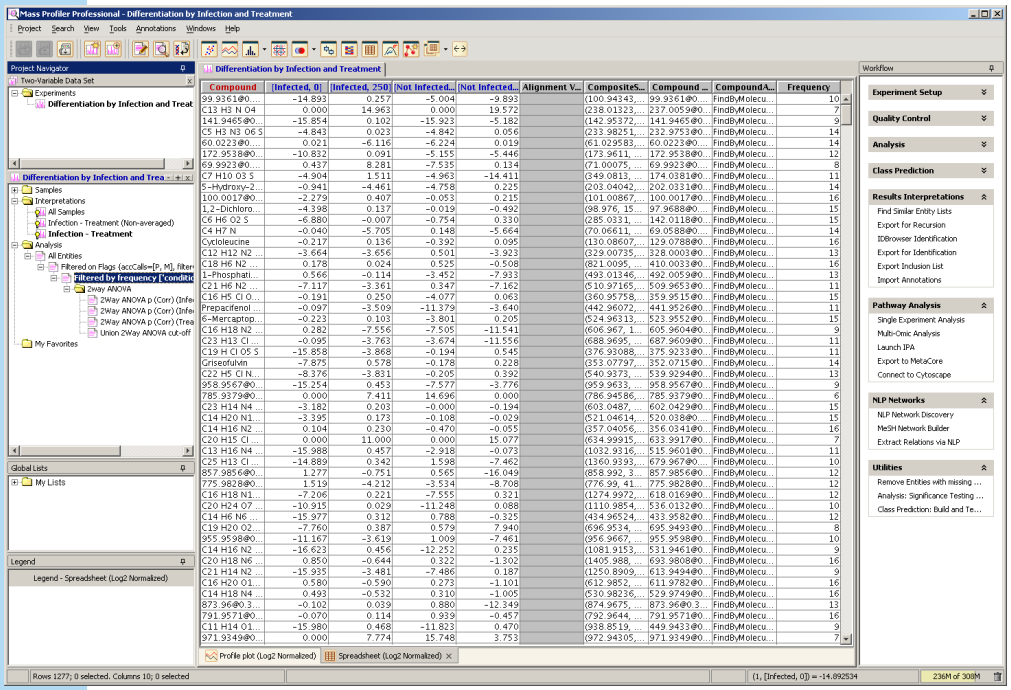


Figure 55 *Spreadsheet view of identified entities in MPP after ID Browser.*

Export for Identification

For an unidentified experiment, this operation allows you to save selected entities for identification with another program. Export the entities in a selected entity list to a CEF file (Compound Exchange Format).

1. Launch **Export for Identification** in the Workflow Browser.

a Click **Export for Identification** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Export for Identification operation has one (1) step as shown in Figure 56.

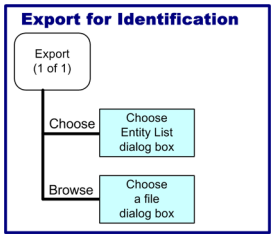


Figure 56 *Flow chart of the Export for Identification operation.*

2. Select the entity list to export.

a Click **Choose** in the **Export** dialog box.

b Select the entity list to export.

Since this is an identification operation, you do not need to select an entity list that is at least **Filtered on Flags** from the entity lists in the **Choose Entity List** dialog box.

c Click **OK**.

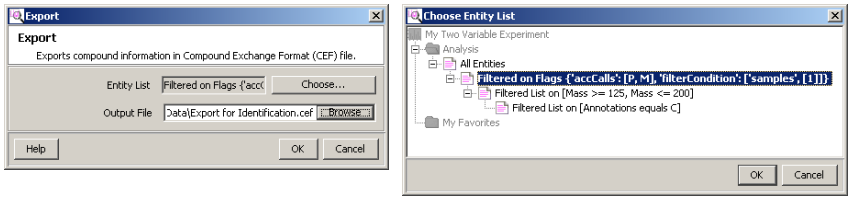


Figure 57 *Export and Choose Entity List dialog boxes*

3. Enter the export file name and folder.

a Click **Browse** in the **Export** dialog box.

Do not type a file name at this location.

b Select the folder or create a new folder for your CEF file in the **Choose a file** dialog box.

c Type the **File name**. For example, you can type `Export for Identification.cef`.

d Click **Save**.

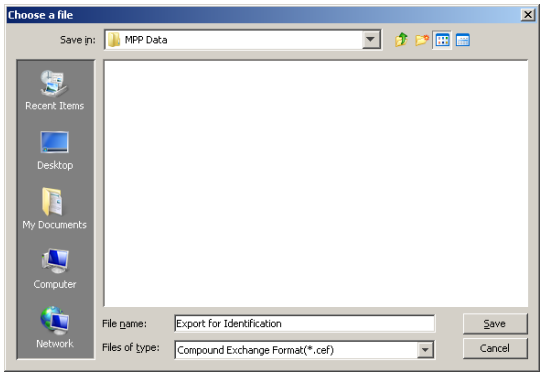


Figure 58 *Choose a file dialog box.*

e Click **OK**.

Export Inclusion List

Export inclusion parameters from the specified entity list. This operation produces a CSV file format (comma separated variable) and is applicable to MassHunter Qualitative Analysis, MassHunter Qualitative Analysis GC Scan, AMDIS, and ChemStation experiment creation.

1. Launch **Export Inclusion List** in the Workflow Browser.

a Click **Export Inclusion List** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Export Inclusion List operation has two (2) steps as shown in Figure 59.

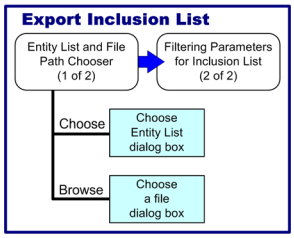

Export Inclusion List
Entity List and File Path Chooser (1 of 2)
Filtering Parameters for Inclusion List (2 of 2)
Choose
Choose Entity List dialog box
Browse
Choose a file dialog box

Figure 59 *Flow chart of the Export Inclusion List operation.*

2. Select the entity list in **Export Inclusion List (Step 1 of 2)**.

a Click **Choose**.

b Select the entity list to export.

Click an entity list that is at least **Filtered on Flags** from the entity lists in the **Choose Entity List** dialog box. More significance in your analysis is obtained by selecting an entity list that has at least been filtered by flags to remove one-hit wonders.

c Click **OK**.

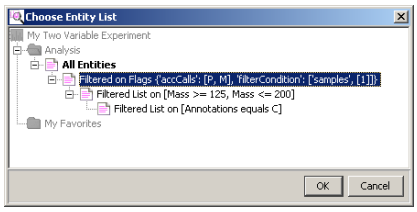


Figure 60 *Choose Entity List dialog box*

d Click **Browse**.

Do not type a file name at this location.

e Select the folder or create a new folder for your CEF file in the **Choose a file** dialog box.

f Type the **File name**. For example, you can type `Export Inclusion List.csv`.

g Click **Save**.

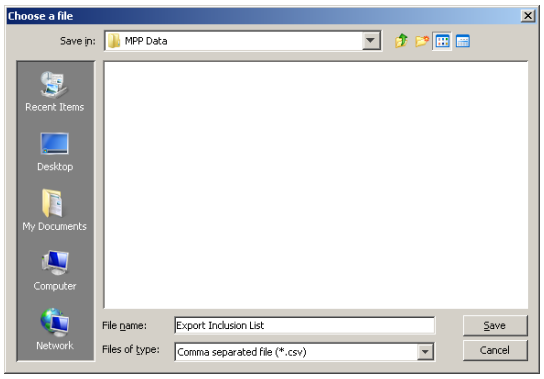


Figure 61 *Choose a file dialog box*

h Click **Next**.

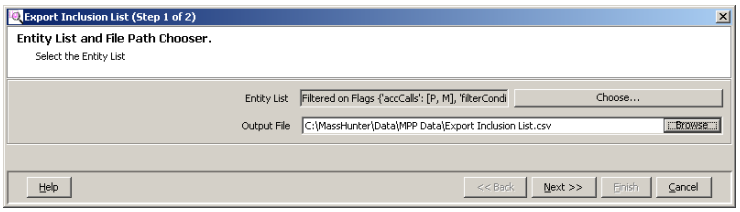


Figure 62 *Entity List and File Path Chooser page (Export Inclusion List (Step 1 of 2))*

3. Enter filter parameters in **Export Inclusion List (Step 2 of 2)**.

a Type values in the **Retention time window**. The default values are 0.0 percent and 0.25 min.

b Mark the **Limit number of precursor ions per compound to** check box and type in a value for **ion(s)** per compound. By default this check box is cleared and the default value is 1 ion.

c Mark the **Minimum ion abundance** and type in the minimum ion **counts**. By default, this check box is cleared and the default value is 2000 counts.

If the sample data is from MassHunter Qualitative Analysis all of the filter options are available. Sample data from other experiment types, non-MassHunter Qualitative Analysis sample data, cannot be processed using the Positive ions, Negative ions, Exported *m/z* value, and Charge state preference filters.

d Select **Export monoisotopic *m/z*** as the monoisotopic value or the value represented by the ion with the highest abundance.

e Select **Specify charge state preference order** to activate the inactive and active charge state options. Specify the highest abundance charge state or as specified by the charge state preference order.

f Mark the **Positive ions** and **Negative ions** that are included in the filter.

g Click **Finish**.

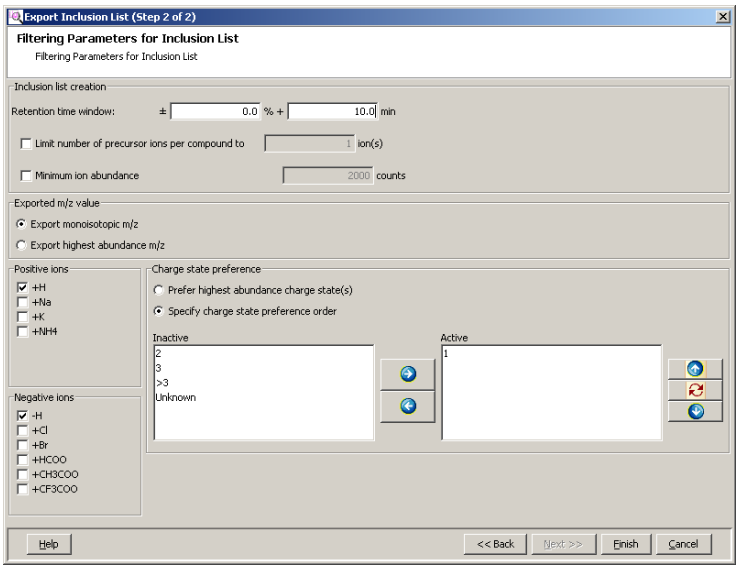


Figure 63 *Filtering Parameters for Inclusion List page (Export Inclusion List (Step 2 of 2))*

Inclusion filter application order

The inclusion filters are applied in following order:

1. **Positive Ions** and **Negative Ions** filters. Peaks which contain the selected ions are passed; e.g. if only +H and +Na ions are marked then peaks with ion species similar to M+H, M+2H, M+Na, M+2Na, ... M+H+Na are selected for further filtering.
2. **Peaks with same charge state and same ion species are grouped in one isotope cluster** (e.g. M+H, M+H+1, M+H+2 in one cluster). From this cluster only one peak is exported depending upon the selection for the **Exported *m/z* value** filter. If **Export monoisotopic m/z** is selected then the peak similar to the M+H (or M+Na, M+2H, etc) is selected from isotope cluster. Otherwise, if the filter **Export highest abundance m/z** is selected then the peak is exported which has the maximum abundance in each isotope cluster.
3. **Charge State Preference** filter. If **Prefer highest abundance charge state(s)** is selected then the peaks per compound are listed in descending order of abundance. Otherwise if **Specify charge state preference order** is selected only those peaks whose charge states are specified in the **Active** window are passed and ordered as specified. For example, if you specified Charge states as 2, 3, and >3 then peaks with charge states 1 are filtered out and the peaks with charge states 2, 3, and >3 are passed. The results are ordered with charge state 2 then all peaks with charge state 3 and finally those with a charge state >3 in descending order of abundance.
4. **Minimum Ion abundance** filter passes only the peak ions with an abundance greater than the specified value.
5. **Limit number of precursor ions** passes only the top number of peaks/compounds as specified.

4. Review the exported inclusion list.

The results from **Export Inclusion List** are saved in a CSV file and include the *m/z*, charge state, retention time, and delta retention time. You can review your results without the Mass Profiler Professional software.

a Open the CSV file using a text editor or spreadsheet program.

b Review the results, see Figure 64.

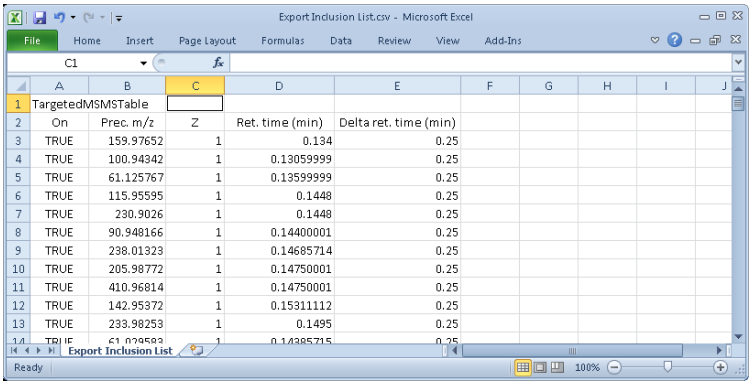


Figure 64 *Contents of an Export Inclusion List CSV file.*

Import Annotations

This operation imports annotations from an identified CEF file and applies the annotations to matching entities in your experiment. When you invoke this operation, you select a CEF file and update annotations for compounds whose Mass Profiler Professional ID match that of the compounds in the imported CEF file. All entity lists in your experiment are updated.

1. Launch **Import Annotations** in the Workflow Browser.

a Click **Import Annotations** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Import Annotations operation has one (1) step as shown in Figure 65.

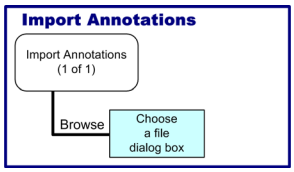


Figure 65 *Flow chart of the Import Annotations operation.*

2. Select the entity list to import annotations.

a Click **Browse**.

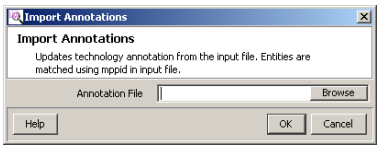


Figure 66 *Import Annotations dialog box*

b Select the folder containing your CEF file in the **Choose a file** dialog box.

c Type or click the **File name**.

d Click **Open**.

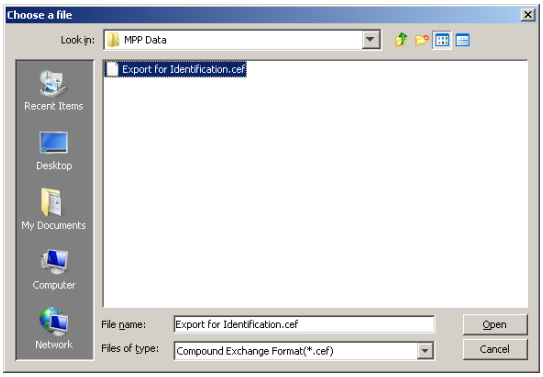


Figure 67 *Choose a file dialog box*

e Click **OK**. A progress box is displayed while the annotations are updated.

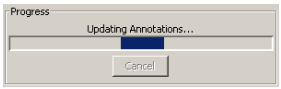
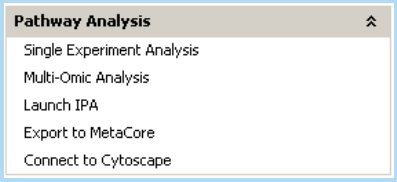


Figure 68 *Progress indication while annotations are updated.*

Pathway Analysis



Analysis of 'omics' data in Mass Profiler Professional typically results in a list of entities that are significantly different in the experimental conditions of interest. Pathway analysis provides the necessary biological context for a functional analysis of these entities to better understand their role in a biological process.

Pathway Analysis supports analysis on well-studied, curated pathways, while the NLP Network Discovery component drives discovery by creating networks around the entities of interest using a powerful Natural Language Processing (NLP) algorithm that extracts information from published literature.

Note: The Pathway Analysis features in Mass Profiler Professional are licensed separately and can only be accessed with a valid Pathway Architect module license. See "Getting started requirements" on page 62.

Pathway Analysis consists of five operations. These operations can only be performed on entity lists that have been annotated, and in some cases on entity lists with Entrez Gene ID annotation. Annotation of your entity list can be done, for example, using ID Browser and SimLipid.

- "Single Experiment Analysis" on page 63
- "Multi-Omic Analysis" on page 71
- "Launch IPA" on page 76
- "Export to MetaCore" on page 82
- "Connect to Cytoscape" on page 85

Features of Pathway Analysis

The following features in Pathway Analysis help you interpret your experiment:

- Import curated pathways directly from WikiPathways portal (http://www.wikipathways.org) and BioCyc (http://www.biocyc.com) or import pathways from other sources in the BioPAX (Level 2 and Level 3), GPML, or Text format.
- Create your own interaction networks from a database of biological and chemical entities, relationships between entities, and properties of these entities and relationships derived from a proprietary Natural Language Processing (NLP) algorithm.
- Determine which of the created or imported pathways have significant overlap with a specified list of entities from one experiment ("Single Experiment Analysis") or two experiments of the same or differing experiment types ("Multi-Omic Analysis").
- View and investigate pathways and interaction networks in an interactive pathway viewer and overlay your experimental data on these pathways.
- Export your data to other popular pathway analysis tools like Ingenuity Pathways Analysis (IPA), MetaCore, and Cytoscape. In the case of IPA, you can also import entity lists resulting from pathway analysis into GeneSpring.

Pathway Analysis can help you answer questions such as:

- What biological pathways and processes are significantly represented by the experiment?
- What other entities and pathways reported in literature are affected by the results of the experiment?
- Is there a pattern in the expression of connected genes across different experimental conditions or is there a pattern of different entity types as measured by experiments under similar conditions? Overlaying data on a signaling pathway can provide an understanding of the cause and effect relationships

between the genes or proteins of interest and provide insight into the mechanism of a specific condition under study.

- Which small molecules might interact with a gene or set of genes?

Getting started requirements

Pathway Analysis requires the following:

1. A valid license for the **Pathway Architect** module. See the *Agilent G3835AA MassHunter Mass Profiler Professional - Quick Start Guide* and the *Mass Profiler Professional User Manual* for information about licenses in Mass Profiler Professional.
2. A valid license for the **GeneSpring GX** module. See the *Agilent G3835AA MassHunter Mass Profiler Professional - Quick Start Guide* and the *Mass Profiler Professional User Manual* for information about licenses in Mass Profiler Professional.
3. Pathways from sources of interest for the organism under study. See section "11.1.2 Importing Pathways into Mass Profiler Professional" in the *Mass Profiler Professional User Manual* for more information about pathway sources and creating pathways.
4. Supporting databases to perform Pathway Analysis for a single organism or across different organisms. Pathway Analysis is supported by organism specific interaction databases, BridgeDb databases, and HomoloGene annotations. See section "11.1.5 Supporting Databases for Pathway Analysis" in the *Mass Profiler Professional User Manual* for more information.

What is BridgeDb?

Pathways acquired from different sources may refer to the same entity using synonymous names and/or identifiers from different biological databases. Incorporating data from multiple databases leads to variations in annotations. BridgeDb (http://www.bridgedb.org) is an identifier mapping framework for bioinformatics applications and provides mapping for the same entity across different biological databases.

Single Experiment Analysis and Multi-Omics Analysis use BridgeDb to search for pathways that match the entities in the your entity list(s). Click **Annotations > Update BridgeDb > From Agilent Server** to update BridgeDb. See section "11.5.2 BridgeDb - ID Mapping" and section "11.1.5 Supporting Databases for Pathway Analysis" in the *Mass Profiler Professional User Manual* for more information.

Improving your Pathway Analysis results

The results from performing a Pathway Analysis are dependent upon (1) the number of annotated entities in your experiment and (2) the quality of the annotations. Entities from proteomics and genomics experiments typically provide greater pathway analysis accuracy (less ambiguity) because these entities are more highly annotated. When you are using entities from metabolomics experiments you can improve your analysis accuracy by (1) using GCMS data that has been identified in a spectral library search, (2) using data from targeted QQQ experiments, and (3) using ID Browser to annotate data from high resolution LCMS experiments. By comparing the pathway results from filtered entity lists (i.e., fold change, K-means clustering) and unfiltered entity lists (not filtered for fold change or significance) lists can also improve your results.

Single Experiment Analysis

Single Experiment Analysis (SEA) identifies pathways that contain entities in common to the entities in the selected entity list for one experiment. The matched entities are highlighted on the pathway. Commonality between a pathway and an entity is determined via the presence of a shared identifier. The operation works with genomics, transcriptomics, proteomics, and metabolomics experiments. Entity lists may contain genes, proteins, or metabolites. Single Experiment Analysis helps you determine in which biological pathways there exists a significant enrichment of compounds of interest based on the input entity list.

You can choose an organism for pathway analysis that differs from the organism associated with your experiment. Curated pathways, such as WikiPathways, BioCyc pathways, and BioPAX pathways, as well as NLP and MeSH created pathways can be individually selected as sources for Pathway Analysis.

Note: Single Experiment Analysis is referred to as Find Significant Pathways in prior versions of Mass Profiler Professional.

1. Launch **Single Experiment Analysis** in the Workflow Browser.

a Click **Single Experiment Analysis** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Single Experiment Analysis operation has four (4) steps as shown in Figure 69. The new SEA pathway list is placed in the Analysis folder within the Experiment Navigator.

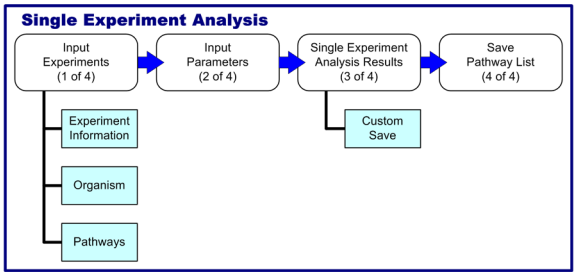


Figure 69 *Flow chart of the Single Experiment Analysis operation.*

2. Select the experiment parameters in **Single Experiment Analysis (Step 1 of 4)**.

a Review the selected **Experiment.** The default experiment is the active experiment in the open project. If available, Click **Choose** to change the **Experiment**.

b Review the **Organism** specified in the **Experiment**. If the specified organism is not specified, or incorrect, you can change the Organism for the experiment in "How to change the organism for an existing experiment".

How to change the organism for an existing experiment

You can change the **Organism** for an experiment:

1. Right-click on the experiment name in the Project Navigator.
2. Click **Inspect Technology**.

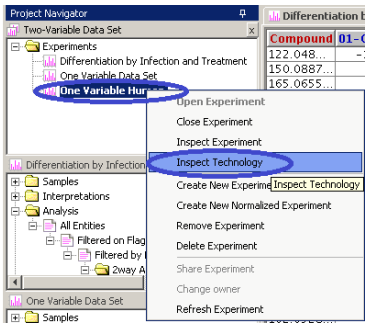


Figure 70 *Inspect Technology for an experiment*

3. Select the **Organism** from the **Technology Inspector** dialog box.

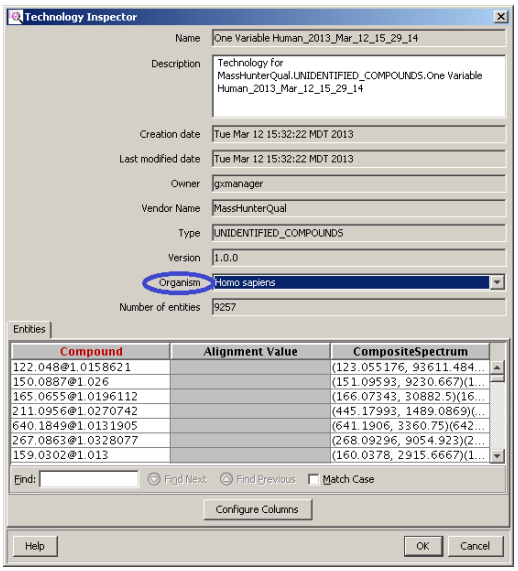


Figure 71 *Inspect Technology for an experiment*

4. Click **OK**.

(End of process to change organism)

c Select the pathway organism in **Choose Pathway Organism**.

You can choose an organism for finding matched pathways that is different from the organism of the selected experiment. Selecting a different organism is useful when the organism specified in the experiment is less or not sufficiently described in the literature, or when you want to observe the effects of one organism's pathogen/metabolite in another organism. By default, the **Choose Pathway Organism** selected is that associated with the **Experiment**.

d Select the pathway source for your analysis.

The following pathway sources are available for **Curated pathways only**:

- WikiPathways - Analysis
- WikiPathways - Reactome
- WikiPathways - GenMAPP
- WikiPathways - Other

- BioCyc/MetaCyc (includes the pathways that you downloaded from the **Agilent Server** using **Tools > Import Pathways from BioCyc**)
- BioPAX (Imported)
- GPML (Imported)
- Hand created
- Legacy

The following pathway sources are available for **Literature Derived Networks only**:

- NLP
- MeSH term
- The pathway sources include interaction networks you imported or created using the **NLP Network Discovery**, **MeSH Network Builder**, or **Extract Relations via NLP** operations in the Workflow Browser.

If you select **Both** then all of the **Curated pathways** and **Literature Derived Networks** pathway sources are available.

e Mark the **Curated pathways** and/or **Literature Derived Networks** to include in your analysis. The number of pathways previously imported into Mass Profiler Professional for each of the sources for the selected pathway organism is displayed in parentheses next to the source name. The number of pathways automatically updates when you choose a different pathway organism for your analysis.

If the number of pathways previously imported into Mass Profiler Professional is reported as zero (0) for your organism among the sources, click **Cancel** and import pathways for your organism. To import pathways for an organism from WikiPathways follow steps in "How to import pathways from WikiPathways" and then return to this step.

How to import pathways from WikiPathways

You can import organism-specific pathways into Mass Profiler Professional from WikiPathways:

1. Click **Tools > Import Pathways from WikiPathways** on the menu bar. A progress status box is displayed while the content is updated.

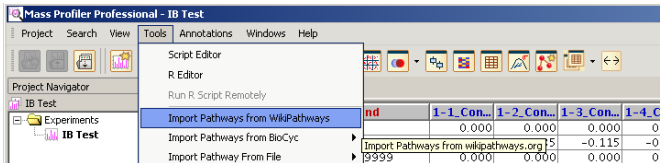


Figure 72 *Importing pathways from WikiPathways*

2. Select **Select Organism** and then select the specific organism from the **Choose Organism** dialog box. Selecting **All Organisms** downloads the pathways for all organisms available in WikiPathways and required additional time to complete. A progress status box is displayed during downloading.

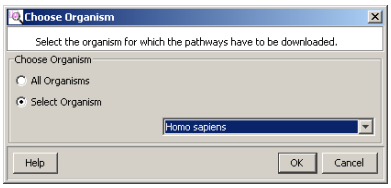


Figure 73 *Choose Organism dialog box*

3. Review the pathways that were imported into Mass Profiler Professional in the **Import Statistics** dialog box.

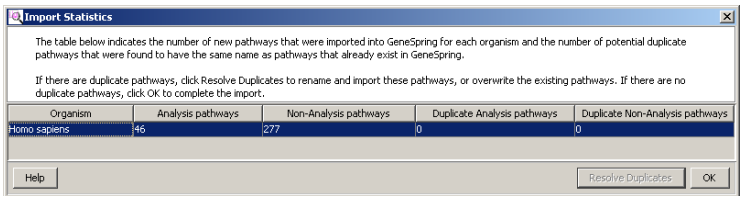

Import Statistics
The table below indicates the number of new pathways that were imported into GeneSpring for each organism and the number of potential duplicate pathways that were found to have the same name as pathways that already exist in GeneSpring.
If there are duplicate pathways, click Resolve Duplicates to rename and import these pathways, or overwrite the existing pathways. If there are no duplicate pathways, click OK to complete the import.
Organism
Analysis pathways
Non-Analysis pathways
Duplicate Analysis pathways
Duplicate Non-Analysis pathways
Homo sapiens
46
277
0
0
Help
Resolve Duplicates
OK

Figure 74 *Import Statistics dialog box*

4. Click **OK**. If BridgeDb databases have not yet been downloaded for the chosen organism, you are prompted with the option to download the corresponding database.

(End of the process to import a pathway)

f Click **Next**. A progress status box is displayed while the pathways are searched based on the organism.

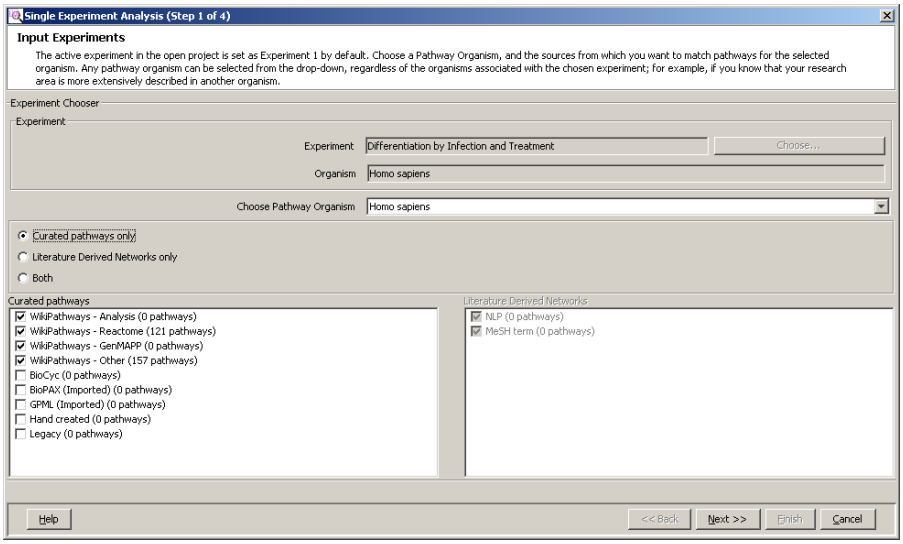

Single Experiment Analysis (Step 1 of 4)
Input Experiments
The active experiment in the open project is set as Experiment 1 by default. Choose a Pathway Organism, and the sources from which you want to match pathways for the selected organism. Any pathway organism can be selected from the drop-down, regardless of the organisms associated with the chosen experiment; for example, if you know that your research area is more extensively described in another organism.
Experiment Chooser
Experiment
Experiment
Differentiation by Infection and Treatment
Choose...
Organism
Homo sapiens
Choose Pathway Organism
Homo sapiens
Curated pathways only
Literature Derived Networks only
Both
Curated pathways
WikiPathways - Analysis (0 pathways)
WikiPathways - Reactome (121 pathways)
WikiPathways - GenMAPP (0 pathways)
WikiPathways - Other (157 pathways)
BioCyc (0 pathways)
BioPAX (Imported) (0 pathways)
GPML (Imported) (0 pathways)
Hand created (0 pathways)
Legacy (0 pathways)
Literature Derived Networks
NLP (0 pathways)
MeSH term (0 pathways)
Help
<< Back
Next >>
Finish
Cancel

Figure 75 *Input Experiments page (Single Experiment Analysis (Step 1 of 4))*

3. Select the interpretation and entity list in **Single Experiment Analysis (Step 2 of 4)**.

a Select an interpretation for **Choose Interpretation**. An interpretation specifies how the samples are grouped based on your experimental conditions.

b Select an entity list for **Choose Entity List**.

c Mark the **Annotations** to use in your analysis. At least one annotation must be marked. Table 4 presents the annotations used by Mass Profiler Professional. If an entity does not have the specified annotation it is not matched.

Table 4 *Annotations used by Mass Profiler Professional*

Available Annotations by "Omic" Type		
Genomics/ Transcriptomics	Proteomics	Metabolomics
RefSeq Protein ID	Swiss-Prot	KEGG ID
RefSeq Transcript ID	NCBI GI number	ChEBI ID
Entrez Gene ID		HMP ID
Ensembl Gene ID		CAS Number
GenBank Accession		
UniGene ID		
TIGR ID		

Entities from the selected entity list and pathways from the selected organism are matched based on their annotation identifiers. If the selected pathway organism differs from the experiment organisms, matching is accomplished by identifying homologous genes based on Entrez Gene IDs using HomoloGene Translation for Gene/Protein Identifiers (http://www.ncbi.nlm.nih.gov/homologene). When the pathway and experiment organism are the same, annotation identifiers are matched using BridgeDb - ID Mapping.

Mass Profiler Professional first tries to find direct matches between the pathway entities and the entities in the selected entity list. A direct match occurs when entities from both the pathways and entity list have identifiers from the same annotation. When identifiers from differing annotations are matched, the BridgeDb algorithm looks for a match in the order in which the annotations are displayed on this wizard page. The first matching annotation and corresponding identifier are displayed in the Heatmap of the Pathway View.

d Click **Next**. A progress status box is displayed while the pathways are searched based on the entities in the entity list.

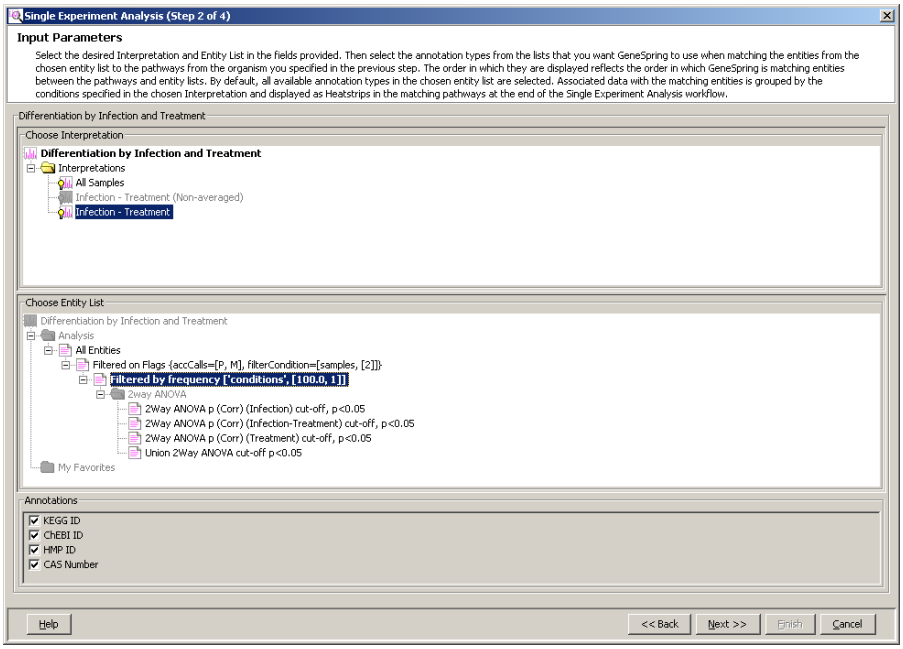

Single Experiment Analysis (Step 2 of 4)
Input Parameters
Select the desired Interpretation and Entity List in the fields provided. Then select the annotation types from the lists that you want GeneSpring to use when matching the entities from the chosen entity list to the pathways from the organism you specified in the previous step. The order in which they are displayed reflects the order in which GeneSpring is matching entities between the pathways and entity lists. By default, all available annotation types in the chosen entity list are selected. Associated data with the matching entities is grouped by the conditions specified in the chosen Interpretation and displayed as Heatstrips in the matching pathways at the end of the Single Experiment Analysis workflow.
Differentiation by Infection and Treatment
Choose Interpretation
Differentiation by Infection and Treatment
Interpretations
All Samples
Infection - Treatment (Non-averaged)
Infection - Treatment
Choose Entity List
Differentiation by Infection and Treatment
Analysis
All Entities
Filtered on Flags {accCalls=[P, M], filterCondition=[samples, [2]]}
Filtered by frequency ['conditions', [100.0, 1]]
2way ANOVA
2Way ANOVA p (Corr) (Infection) cut-off, p<0.05
2Way ANOVA p (Corr) (Infection-Treatment) cut-off, p<0.05
2Way ANOVA p (Corr) (Treatment) cut-off, p<0.05
Union 2Way ANOVA cut-off p<0.05
My Favorites
Annotations
KEGG ID
ChEBI ID
HMP ID
CAS Number
Help
<< Back
Next >>
Finish
Cancel

Figure 76 *Input Parameters page (Single Experiment Analysis (Step 2 of 4))*

4. Review analysis results in **Single Experiment Analysis (Step 3 of 4)**.

Save a custom pathway list

a Review your pathway results.

b *(Optional)* Select one or more pathways to save them as a custom pathway list.

1. Click one or more pathways. See "Review the search results in EntityList Search Wizard (Step 2 of 2)." on page 41 for selecting multiple rows.
2. Click **Custom Save**. This option is only available if one or more pathways are selected; a selected pathway row is highlighted.

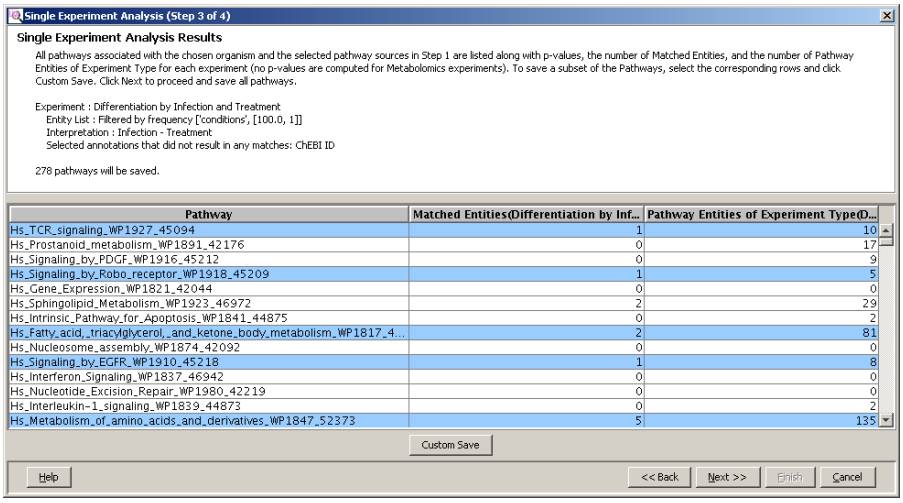

Single Experiment Analysis (Step 3 of 4)
Single Experiment Analysis Results
All pathways associated with the chosen organism and the selected pathway sources in Step 1 are listed along with p-values, the number of Matched Entities, and the number of Pathway Entities of Experiment Type for each experiment (no p-values are computed for Metabolomics experiments). To save a subset of the Pathways, select the corresponding rows and click Custom Save. Click Next to proceed and save all pathways.
Experiment : Differentiation by Infection and Treatment
Entity List : Filtered by frequency ['conditions', [100.0, 1]]
Interpretation : Infection - Treatment
Selected annotations that did not result in any matches: ChEBI ID
278 pathways will be saved.
Pathway
Matched Entities(Differentiation by Inf...
Pathway Entities of Experiment Type(D...
Hs_TCR_signaling_WP1927_45094 1 10
Hs_Prostanoid_metabolism_WP1891_42176 0 17
Hs_Signaling_by_PDGF_WP1916_45212 0 9
Hs_Signaling_by_Robo_receptor_WP1918_45209 1 5
Hs_Gene_Expression_WP1821_42044 0 0
Hs_Sphingolipid_Metabolism_WP1923_46972 2 29
Hs_Intrinsic_Pathway_for_Apoptosis_WP1841_44875 0 2
Hs_Fatty_acid,_triacylglycerol,_and_ketone_body_metabolism_WP1817_4... 2 81
Hs_Nucleosome_assembly_WP1874_42092 0 0
Hs_Signaling_by_EGFR_WP1910_45218 1 8
Hs_Interferon_Signaling_WP1837_46942 0 0
Hs_Nucleotide_Excision_Repair_WP1980_42219 0 0
Hs_Interleukin-1_signaling_WP1839_44873 0 2
Hs_Metabolism_of_amino_acids_and_derivatives_WP1847_52373 5 135
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 77 *Pathways selection on the Single Experiment Analysis Results page (Single Experiment Analysis (Step 3 of 4))*

3. Review the content and parameters in the **Pathway List Inspector** dialog box.

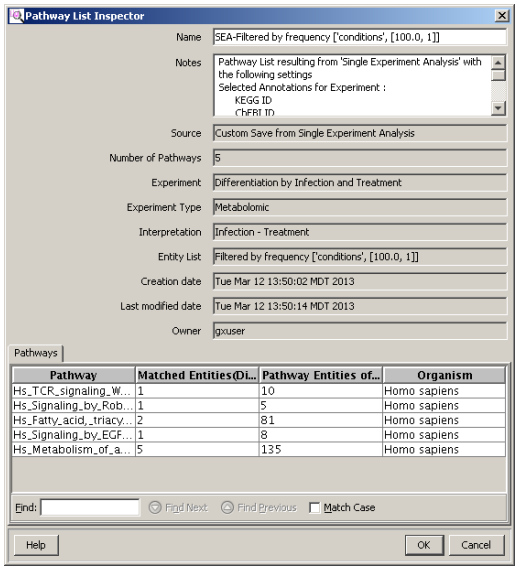


Figure 78 *Pathway List Inspector dialog box*

4. Add or edit descriptive information that is stored with the saved pathway list in the **Name** and **Notes** fields.
5. Click **OK**. The new pathway list is placed in the Analysis folder within the Experiment Navigator.

(End of the optional procedure to select one or more pathways to save them as a custom pathway list)

c Click **Next**.

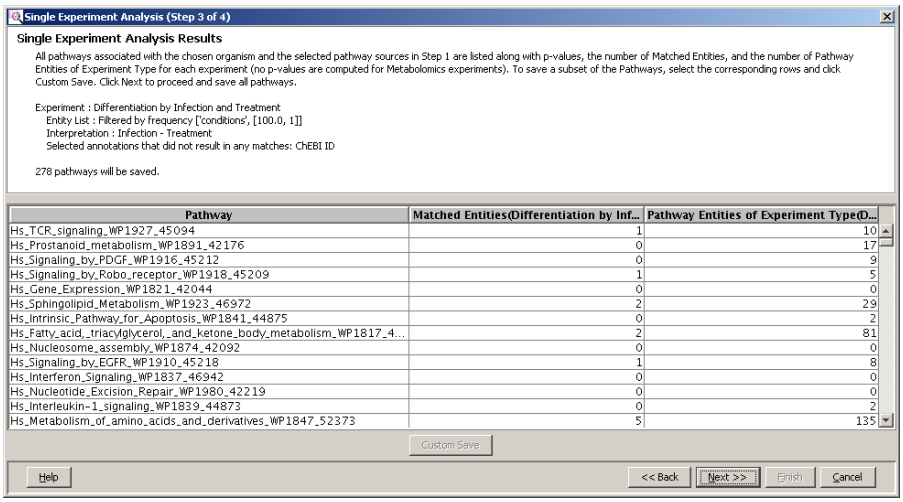

Single Experiment Analysis (Step 3 of 4)
Single Experiment Analysis Results
All pathways associated with the chosen organism and the selected pathway sources in Step 1 are listed along with p-values, the number of Matched Entities, and the number of Pathway Entities of Experiment Type for each experiment (no p-values are computed for Metabolomics experiments). To save a subset of the Pathways, select the corresponding rows and click Custom Save. Click Next to proceed and save all pathways.
Experiment : Differentiation by Infection and Treatment
Entity List : Filtered by frequency ['conditions', [100.0, 1]]
Interpretation : Infection - Treatment
Selected annotations that did not result in any matches: ChEBI ID
278 pathways will be saved.
Pathway
Matched Entities(Differentiation by Inf...
Pathway Entities of Experiment Type(D...
Hs_TCR_signaling_WP1927_45094 1 10
Hs_Prostanoid_metabolism_WP1891_42176 0 17
Hs_Signaling_by_PDGF_WP1916_45212 0 9
Hs_Signaling_by_Robo_receptor_WP1918_45209 1 5
Hs_Gene_Expression_WP1821_42044 0 0
Hs_Sphingolipid_Metabolism_WP1923_46972 2 29
Hs_Intrinsic_Pathway_for_Apoptosis_WP1841_44875 0 2
Hs_Fatty_acid,_triacylglycerol,_and_ketone_body_metabolism_WP1817_4... 2 81
Hs_Nucleosome_assembly_WP1874_42092 0 0
Hs_Signaling_by_EGFR_WP1910_45218 1 8
Hs_Interferon_Signaling_WP1837_46942 0 0
Hs_Nucleotide_Excision_Repair_WP1980_42219 0 0
Hs_Interleukin-1_signaling_WP1839_44873 0 2
Hs_Metabolism_of_amino_acids_and_derivatives_WP1847_52373 5 135
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 79 *Single Experiment Analysis Results page (Single Experiment Analysis (Step 3 of 4))*

5. Enter save pathway list parameters in **Single Experiment Analysis (Step 4 of 4)**.

a Review your pathway list results.

b Add or edit descriptive information that is stored with the saved pathway list in the **Name** and **Notes** fields.

c Click **Next**. The new SEA pathway list is placed in the Analysis folder within the Experiment Navigator.

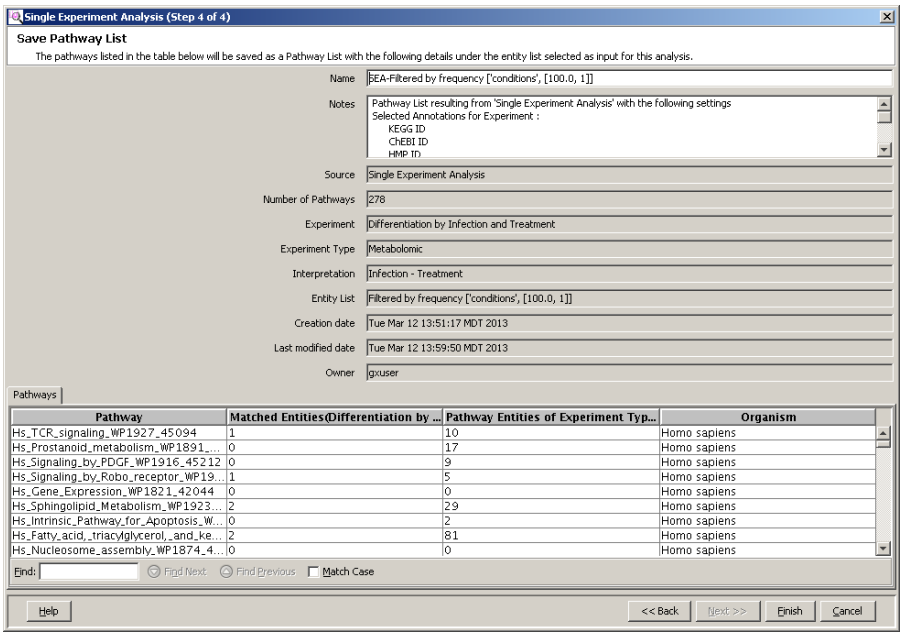


Figure 80 *Save Pathway List page (Single Experiment Analysis (Step 4 of 4))*

The Mass Profiler Professional Display Plane returns showing your entities and associated pathways in the Pathway View as shown in Figure 81. See section "11.3.5 Working with Pathway Lists" in the *Mass Profiler Professional User Manual* for more information about navigating the Pathway View.

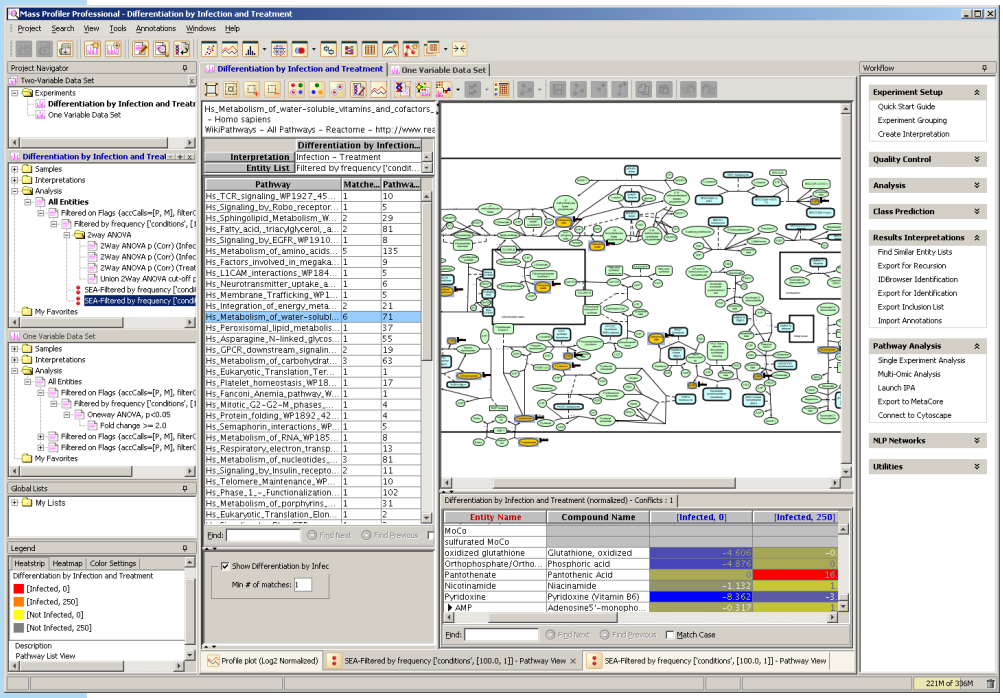


Figure 81 *Pathway View after a Single Experiment Analysis.*

Multi-Omic Analysis

Multi-Omic Analysis (MOA) compares two experiments, and for non-metabolomics experiments has options for you to isolate significant pathways based on the p-value cut-off and the minimum number of matched entities.

With Multi-Omic Analysis you can overlay data from two different experiments on the same pathway, thus performing a simultaneous integrated analysis of data from different experiment types. You can choose an organism for pathway analysis that differs from the organism associated with your experiment and identify significant pathways for data from any combination of genomics, transcriptomics, proteomics, and metabolomics experiments.

This operation finds all pathways that contain entities in common to the entities in the selected entity lists. Commonness between a pathway and an entity is determined via the presence of a shared identifier. The operation works with genomics, transcriptomics, proteomics, and metabolomics experiments. Entity lists may contain genes, proteins, or metabolites. Multi-Omics Analysis helps you determine in which biological pathways there exists a significant enrichment of compounds of interest based on the input entity list.

You can choose an organism for pathway analysis that differs from the organism associated with your experiment. Curated pathways, such as WikiPathways, BioCyc pathways, and BioPAX pathways, as well as NLP and MeSH created pathways can be individually selected as sources for Pathway Analysis.

1. Launch **Multi-Omics Analysis** in the Workflow Browser.

a Click **Multi-Omics Analysis** in the Workflow Browser.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

The Multi-Omics Analysis operation has four (4) steps as shown in Figure 82. The MOA results are assigned a new project in the Project navigator and the MOA pathway lists are placed in the Analysis folder within the Experiment Navigator.

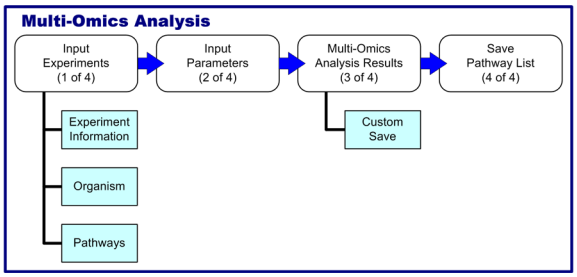


Figure 82 *Flow chart of the Multi-Omics Analysis operation.*

2. Select the experiment parameters in **Multi-Omics Analysis (Step 1 of 4)**.

a Review the selected **Experiment 1.** The default experiment is the active experiment in the open project. If available, Click **Choose** to change the **Experiment**.

b Click **Choose** to select **Experiment 2**. The experiment selected for **Experiment 2** must be different from the experiment selected for **Experiment 1**.

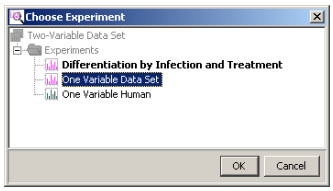


Figure 83 *Choose Experiment dialog box*

c Review the **Organism** specified in **Experiment 1** and **Experiment 2**. If the specified organism is not specified, or incorrect, you can change the Organism for the experiment in "How to change the organism for an existing experiment" on page 63.

d Select the pathway organism in **Choose Pathway Organism**.

You can choose an organism for finding matched pathways that is different from the organism of the selected experiments. Selecting a different organism is useful when the organism specified in the experiment is less or not sufficiently described in the literature, or when you want to observe the effects of one organism's pathogen/metabolite in another organism. By default, the **Choose Pathway Organism** selected is that associated with the **Experiment**.

e Select the pathway source for your analysis.

The following pathway sources are available for **Curated pathways only**:
- WikiPathways - Analysis
- WikiPathways - Reactome
- WikiPathways - GenMAPP
- WikiPathways - Other
- BioCyc/MetaCyc (includes the pathways that you downloaded from the **Agilent Server** using **Tools > Import Pathways from BioCyc**)
- BioPAX (Imported)
- GPML (Imported)
- Hand created
- Legacy

The following pathway sources are available for **Literature Derived Networks only**:
- NLP
- MeSH term
- The pathway sources include interaction networks you imported or created using the **NLP Network Discovery**, **MeSH Network Builder**, or **Extract Relations via NLP** operations in the Workflow Browser.

If you select **Both** then all of the **Curated pathways** and **Literature Derived Networks** pathway sources are available.

f Mark the **Curated pathways** and/or **Literature Derived Networks** to include in your analysis. The number of pathways previously imported into Mass Profiler Professional for each of the sources for the selected pathway organism is displayed in parentheses next to the source name. The number of pathways automatically updates when you choose a different pathway organism for your analysis.

If the number of pathways previously imported into Mass Profiler Professional is reported as zero (0) for your organism among the sources, click **Cancel** and

import pathways for your organism. To import pathways for an organism from WikiPathways follow steps in "How to import pathways from WikiPathways" on page 65 and then return to this step.

g Click **Next**. A progress status box is displayed while the pathways are searched based on the organism.

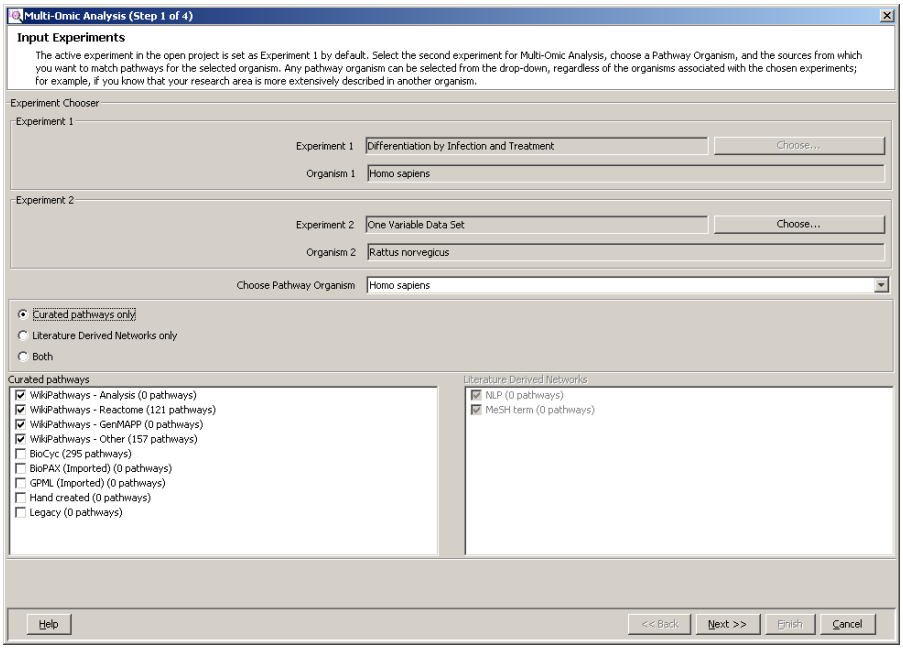

Multi-Omic Analysis (Step 1 of 4)
Input Experiments
The active experiment in the open project is set as Experiment 1 by default. Select the second experiment for Multi-Omic Analysis, choose a Pathway Organism, and the sources from which you want to match pathways for the selected organism. Any pathway organism can be selected from the drop-down, regardless of the organisms associated with the chosen experiments; for example, if you know that your research area is more extensively described in another organism.
Experiment Chooser
Experiment 1
Experiment 1
Differentiation by Infection and Treatment
Choose...
Organism 1
Homo sapiens
Experiment 2
Experiment 2
One Variable Data Set
Choose...
Organism 2
Rattus norvegicus
Choose Pathway Organism
Homo sapiens
Curated pathways only
Literature Derived Networks only
Both
Curated pathways
WikiPathways - Analysis (0 pathways)
WikiPathways - Reactome (121 pathways)
WikiPathways - GenMAPP (0 pathways)
WikiPathways - Other (157 pathways)
BioCyc (295 pathways)
BioPAX (Imported) (0 pathways)
GPML (Imported) (0 pathways)
Hand created (0 pathways)
Legacy (0 pathways)
Literature Derived Networks
NLP (0 pathways)
MeSH term (0 pathways)
Help
<< Back
Next >>
Finish
Cancel

Figure 84 *Input Experiments page (Multi-Omic Analysis (Step 1 of 4))*

3. Select the interpretation and entity list in **Multi-Omic Analysis (Step 2 of 4)**.

a Select an interpretation for **Choose Interpretation** for each experiment. An interpretation specifies how the samples are grouped based on your experimental conditions.

b Select an entity list for **Choose Entity List** for each experiment.

c Select **Annotations** for each experiment to use in your analysis. At least one annotation must be specified. Table 4 on page 67 presents the annotations used by Mass Profiler Professional.

Entities from the selected entity list and pathways from the selected organism are matched based on their annotation identifiers. If the selected pathway organism differs from the experiment organisms, matching is accomplished by identifying homologous genes based on Entrez Gene IDs using HomoloGene Translation for Gene/Protein Identifiers (http://www.ncbi.nlm.nih.gov/homologene). When the pathway and experiment organism are the same, annotation identifiers are matched using BridgeDb - ID Mapping.

Mass Profiler Professional first tries to find direct matches between the pathway entities and the entities in the selected entity list. A direct match occurs when entities from both the pathways and entity list have identifiers from the same annotation. When identifiers from differing annotations are matched, the BridgeDb algorithm looks for a match in the order in which the annotations are displayed on this wizard page. The first matching annotation and corresponding identifier are displayed in the Heatmap of the Pathway View.

d Click **Next**. A progress status box is displayed while the pathways are searched based on the entities in the entity list.

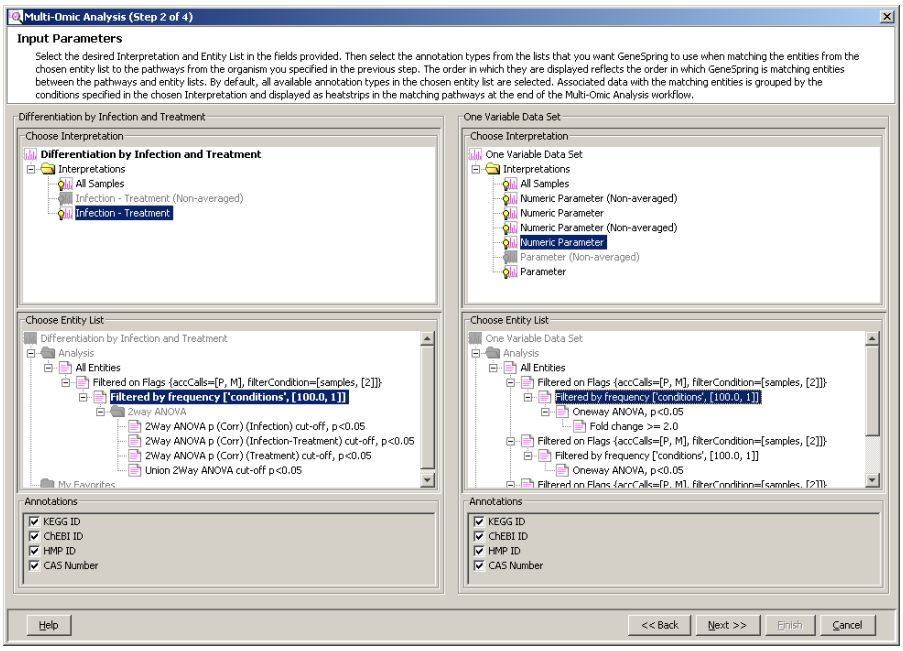

Multi-Omic Analysis (Step 2 of 4)
Input Parameters
Select the desired Interpretation and Entity List in the fields provided. Then select the annotation types from the lists that you want GeneSpring to use when matching the entities from the chosen entity list to the pathways from the organism you specified in the previous step. The order in which they are displayed reflects the order in which GeneSpring is matching entities between the pathways and entity lists. By default, all available annotation types in the chosen entity list are selected. Associated data with the matching entities is grouped by the conditions specified in the chosen Interpretation and displayed as heatstrips in the matching pathways at the end of the Multi-Omic Analysis workflow.
Differentiation by Infection and Treatment
Choose Interpretation
Differentiation by Infection and Treatment
Interpretations
All Samples
Infection - Treatment (Non-averaged)
Infection - Treatment
One Variable Data Set
Choose Interpretation
One Variable Data Set
Interpretations
All Samples
Numeric Parameter (Non-averaged)
Numeric Parameter
Numeric Parameter (Non-averaged)
Numeric Parameter
Parameter (Non-averaged)
Parameter
Choose Entity List
Differentiation by Infection and Treatment
Analysis
All Entities
Filtered on Flags {accCalls=[P, M], filterCondition=[samples, [2]]}
Filtered by frequency ['conditions', [100.0, 1]]
2way ANOVA
2Way ANOVA p (Corr) (Infection) cut-off, p<0.05
2Way ANOVA p (Corr) (Infection-Treatment) cut-off, p<0.05
2Way ANOVA p (Corr) (Treatment) cut-off, p<0.05
Union 2Way ANOVA cut-off p<0.05
My Favorites
Choose Entity List
One Variable Data Set
Analysis
All Entities
Filtered on Flags {accCalls=[P, M], filterCondition=[samples, [2]]}
Filtered by frequency ['conditions', [100.0, 1]]
Oneway ANOVA, p<0.05
Fold change >= 2.0
Filtered on Flags {accCalls=[P, M], filterCondition=[samples, [2]]}
Filtered by frequency ['conditions', [100.0, 1]]
Oneway ANOVA, p<0.05
Annotations
KEGG ID
ChEBI ID
HMP ID
CAS Number
Annotations
KEGG ID
ChEBI ID
HMP ID
CAS Number
Help
<< Back
Next >>
Finish
Cancel

Figure 85 *Input Parameters page (Multi-Omic Analysis (Step 2 of 4))*

4. Review analysis results in **Multi-Omic Analysis (Step 3 of 4)**.

a Review your pathway results.

b *(Optional)* Select one or more pathways to save them as a custom pathways list. See "Review analysis results in Single Experiment Analysis (Step 3 of 4)." on page 68 for the steps involved in saving a custom pathways list.

c Click **Next**.

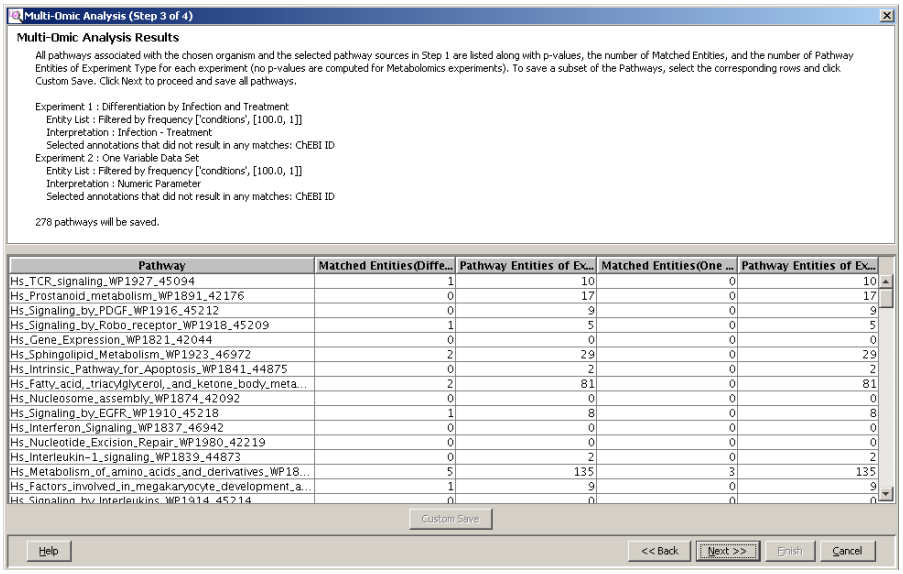

Multi-Omic Analysis (Step 3 of 4)
Multi-Omic Analysis Results
All pathways associated with the chosen organism and the selected pathway sources in Step 1 are listed along with p-values, the number of Matched Entities, and the number of Pathway Entities of Experiment Type for each experiment (no p-values are computed for Metabolomics experiments). To save a subset of the Pathways, select the corresponding rows and click Custom Save. Click Next to proceed and save all pathways.
Experiment 1 : Differentiation by Infection and Treatment
Entity List : Filtered by frequency ['conditions', [100.0, 1]]
Interpretation : Infection - Treatment
Selected annotations that did not result in any matches: ChEBI ID
Experiment 2 : One Variable Data Set
Entity List : Filtered by frequency ['conditions', [100.0, 1]]
Interpretation : Numeric Parameter
Selected annotations that did not result in any matches: ChEBI ID
278 pathways will be saved.
Pathway
Matched Entities(Diffe...
Pathway Entities of Ex...
Matched Entities(One ...
Pathway Entities of Ex...
Hs_TCR_signaling_WP1927_45094 1 10 0 10
Hs_Prostanoid_metabolism_WP1891_42176 0 17 0 17
Hs_Signaling_by_PDGF_WP1916_45212 0 9 0 9
Hs_Signaling_by_Robo_receptor_WP1918_45209 1 5 0 5
Hs_Gene_Expression_WP1821_42044 0 0 0 0
Hs_Sphingolipid_Metabolism_WP1923_46972 2 29 0 29
Hs_Intrinsic_Pathway_for_Apoptosis_WP1841_44875 0 2 0 2
Hs_Fatty_acid,_triacylglycerol,_and_ketone_body_meta... 2 81 0 81
Hs_Nucleosome_assembly_WP1874_42092 0 0 0 0
Hs_Signaling_by_EGFR_WP1910_45218 1 8 0 8
Hs_Interferon_Signaling_WP1837_46942 0 0 0 0
Hs_Nucleotide_Excision_Repair_WP1980_42219 0 0 0 0
Hs_Interleukin-1_signaling_WP1839_44873 0 2 0 2
Hs_Metabolism_of_amino_acids_and_derivatives_WP18... 5 135 3 135
Hs_Factors_involved_in_megakaryocyte_development_a... 1 9 0 9
Hs_Signaling_by_Interleukins_WP1914_45214 0 0 0 0
Custom Save
Help
<< Back
Next >>
Finish
Cancel

Figure 86 *Multi-Omic Results page (Multi-Omic Analysis (Step 3 of 4))*

5. Enter save pathway list parameters in **Multi-Omic Analysis (Step 4 of 4)**.

a Review your pathway list results.

b Add or edit descriptive information that is stored with the saved pathway list in the **Name** and **Notes** fields.

c Click **Next**. The MOA results are assigned a new project in the Project Navigator and the MOA pathway lists are placed in the Analysis folder within the Experiment Navigator.

The Mass Profiler Professional Display Plane returns showing your entities and associated pathways in the Pathway View as shown in Figure 88 on page 76. See section "11.3.5 Working with Pathway Lists" in the *Mass Profiler Professional User Manual* for more information about navigating the Pathway View.

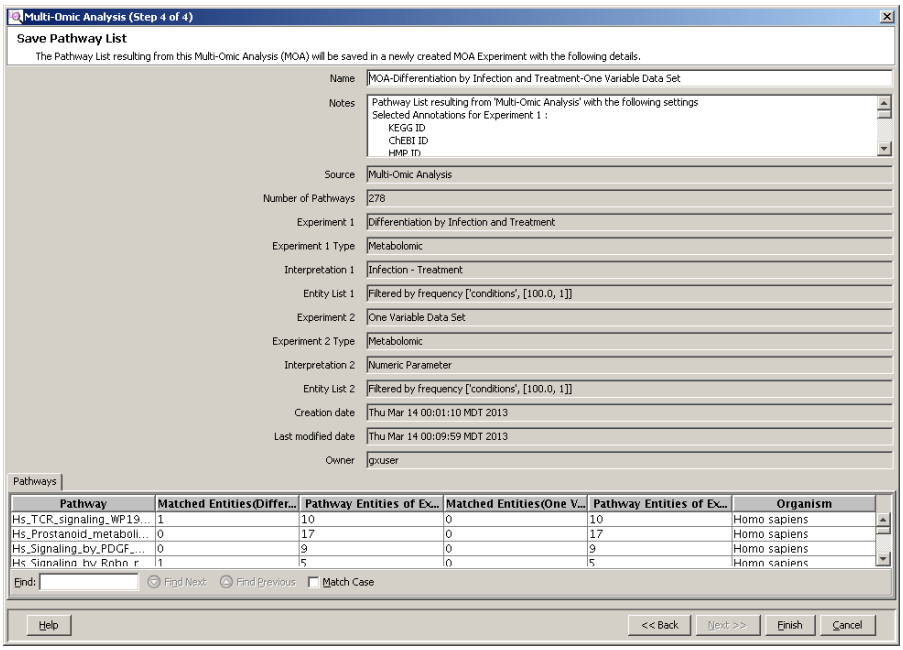


Figure 87 *Save Pathway List page (Multi-Omic Analysis (Step 4 of 4))*

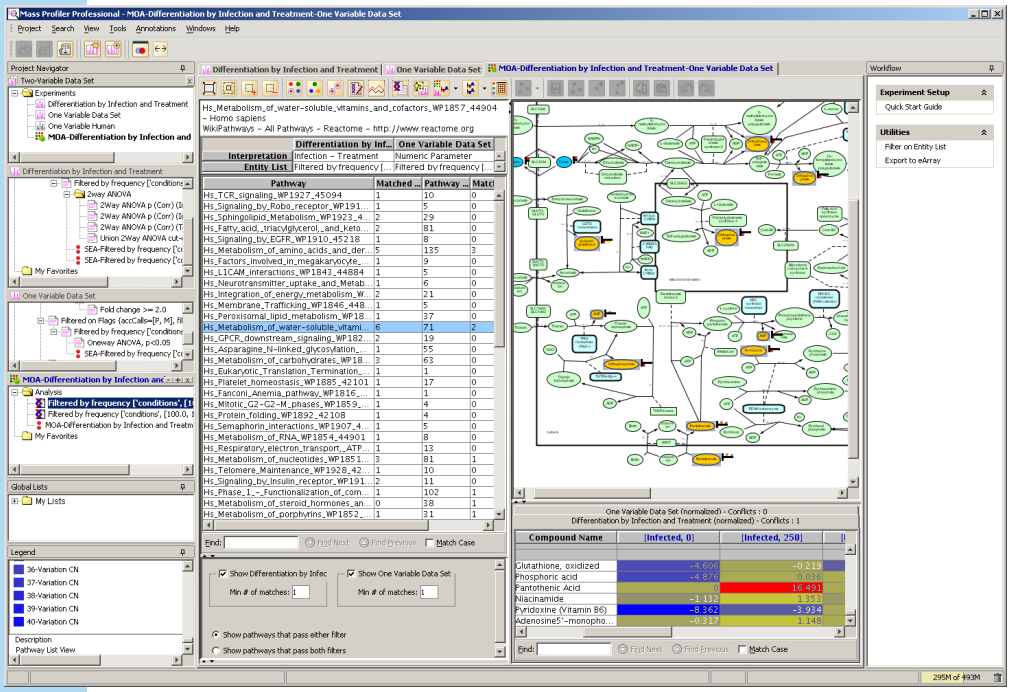


Figure 88 *Pathway View after a Multi-Omics Analysis.*

Launch IPA

Launch IPA enables pathway information exchange between Mass Profiler Professional and Ingenuity Pathways Analysis (IPA, Ingenuity® Systems, www.ingenuity.com). Genes of interest identified using Mass Profiler Professional can be assessed in IPA using its various analysis tools. An IPA account is required to use this operation.

Launch IPA sends gene lists and associated expression data directly to IPA. IPA provides an interface for you to perform network analyses, build pathways, view relevant canonical pathways, and obtain proprietary information on protein interactions and pathways. The IPA interface can send a list of genes back to Mass Profiler Professional (only available for some experiment types), allowing further iterative analysis of those genes.

Note: You must have an account with Ingenuity® Systems (www.ingenuity.com) in order to make use of the Launch IPA operation.

1. Launch **Launch IPA** in the Workflow Browser.

a Click **Launch IPA** in the Workflow Browser.

The Launch IPA operation has one (1) step as shown in Figure 89 on page 77.

This operation is illustrated with data from the "Two-variable experiment" to provide an overview of the wizard options. The data is initially imported and analyzed following the *Agilent Metabolomics Workflow - Discovery Workflow Guide*.

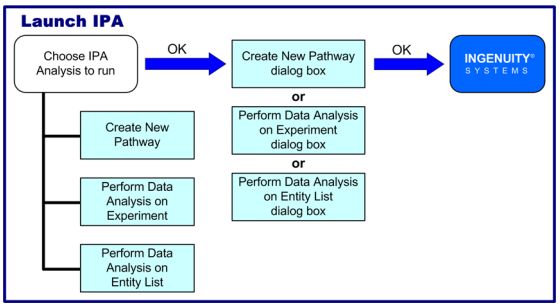


Figure 89 *Flow chart of the Launch IPA operation.*

2. Select the IPA Analysis to run.

a Select the **Choose IPA Analysis to run**.

Create Pathway in IPA sends an Entity List from Mass Profiler Professional to IPA and uses those genes to create a pathway in IPA. This pathway can then be subjected to further manipulation and analysis in IPA by growing a node, removing nodes and interactions, and interrogating a node or an interaction.

Perform Data Analysis on Experiment sends an entity list and the associated gene expression data to IPA to perform data analysis in IPA. Genes in the entity list that are also found in Ingenuity Pathways Knowledge Base (IPKB) are used as Focus Genes to build networks. The networks can be subjected to further manipulation and analysis in IPA by growing a node, removing nodes and interactions, interrogating a node or an interaction, and performing Function, Canonical Pathways, My Pathways, Gene Summary, and Overlapping Networks analyses. You can create gene lists from the generated networks and send the gene lists back to Mass Profiler Professional.

Perform Data Analysis on Entity List sends an entity list, with or without list-associated values, to IPA to perform data analysis in IPA. Genes in the entity list that are also found in IPKB are used as Focus Genes to build networks. The networks can be subjected to further manipulation and analysis in IPA by growing a node, removing nodes and interactions, interrogating a node or an interaction, and perform Function, Canonical Pathways, My Pathways, Gene Summary, and Overlapping Networks analyses. You can create gene lists from the generated networks and send the gene lists back to Mass Profiler Professional.

b Click **OK**. The next step depends on your selection: go to "Enter the options for Create New Pathway." on page 78, "Enter the options for Perform Data Analysis on Experiment." on page 79, or "Enter the options for Perform Data Analysis on Entity List." on page 81.

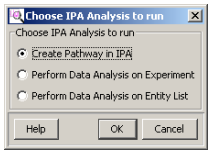


Figure 90 *Choose Experiment dialog box*

3. Enter the options for **Create New Pathway**.

a Click **Choose** to select the **Entity List**. By default, the active entity list is already selected (Figure 91).

b Click **OK**.

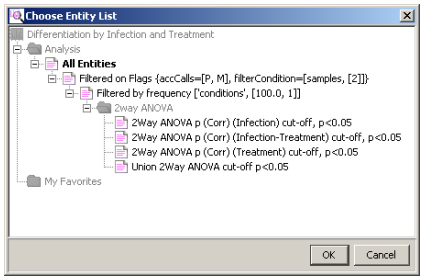

Choose Entity List
Differentiation by Infection and Treatment
Analysis
All Entities
Filtered on Flags {accCalls=[P, M], filterCondition=[samples, [2]]}
Filtered by frequency ['conditions', [100.0, 1]]
2way ANOVA
2Way ANOVA p (Corr) (Infection) cut-off, p<0.05
2Way ANOVA p (Corr) (Infection-Treatment) cut-off, p<0.05
2Way ANOVA p (Corr) (Treatment) cut-off, p<0.05
Union 2Way ANOVA cut-off p<0.05
My Favorites
OK
Cancel

Figure 91 *Choose Entity List dialog box*

c Review the **IPA Server Address**. Type in the address for the IPA server, for example, `analysis.ingenuity.com`.

d Type the name for the new pathway to be created in **Pathway Name**. By default, the name of the entity list that was originally selected is used. If you selected a different entity list above the name for the pathway is not updated to reflect the new entity list selection.

e Type the name of the **Project Folder** that is used by IPA for your analysis. The default name is the same used by Mass Profiler Professional.

f Select the **Gene Identifier Column**. The gene identifier is used to map genes in the entity list to genes in the IPKB.

g *(Optional)* Mark **Save Pathway**. The new pathway is saved in IPA to the specified **Project Folder**, within My Pathways, under the specified **Pathway Name**.

h Click **OK**. Your default Internet browser is automatically launched and connected to the IPA server as specified in the **IPA Server Address**.

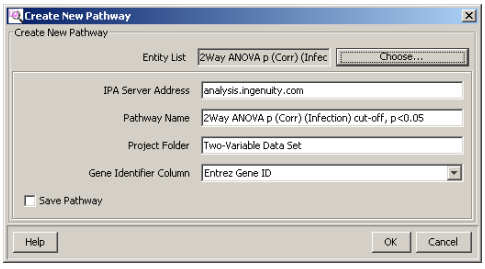


Figure 92 *Create New Pathway dialog box*

i Sign in to IPA as shown in Figure 93 on page 79.

Note: Information on how to use IPA is covered in section "11.6.1 Ingenuity Pathways Analysis (IPA) Connector" in the *Mass Profiler Professional User Manual* and accessed from the Quick Start page of IPA as shown in Figure 94 on page 79.

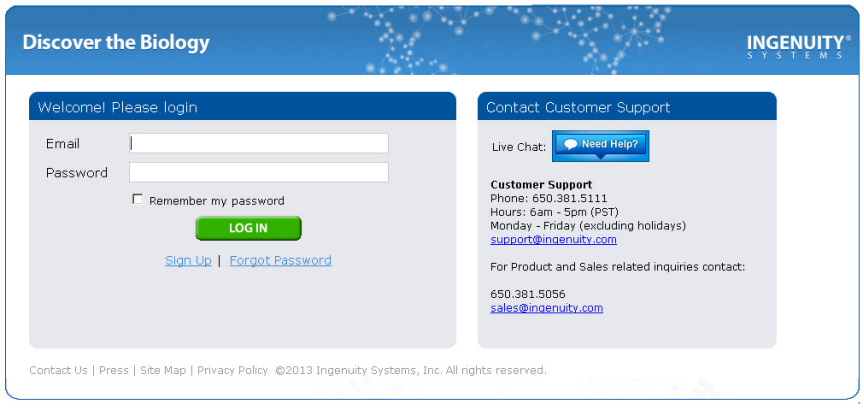


Figure 93 *IPA sign in page*

Quick Start
IPA
Start Here
Learning IPA
Shortcuts
Learn IPA through the easy tutorials below. Each will be checked as you complete it.
View Pathways
Search for Genes
See an Analysis Example
Upload & Analyze Example Data
Upload & Analyze Your Own Expression Data
Visualize Connections Among Genes
Learn about Specialized Features
Explore the Pathway Library
View specific examples or browse the library
View Apoptosis Signaling
View Fatty Acid β-oxidation I
› More Signaling Pathways
› More Metabolic Pathways
For further exploration
How do I open the pathways?
Short Videos about Pathways
What do the symbols and shapes mean?
What do the connections mean?
How can I export or save a copy of these images?
How do I visually relate this to my own data?
Do Not Show at Startup | Tutorials and Help | Training Schedule | Customer Support | How to Cite IPA
CLOSE

Figure 94 *IPA Quick Start page*

4. Enter the options for **Perform Data Analysis on Experiment**.

a Review the **Entity List**. The active entity list is selected. To use a different entity list, cancel the operation, select a different entity list in the Experiment Navigator, and relaunch the operation.

b Click **Choose** to select the **Experiment Interpretation**. By default, the active interpretation is already selected (Figure 95).

Log_2 values for the conditions in the selected experiment interpretation are sent to IPA for analysis. The name of the data set used in IPA is named after the source experiment in Mass Profiler Professional.

c Click **OK**.

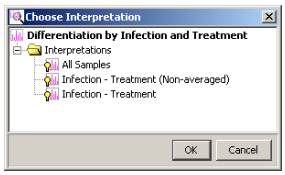


Figure 95 *Choose Interpretation dialog box*

d Review the **IPA Server Address**. Type in the address for the IPA server, for example, `analysis.ingenuity.com`.

e Type the name for the project to be created in **Project Name**. By default, the name used for the experiment that was originally selected is used. The **Project Name** is used by IPA under to store the pathway information.

Note: IPA only allows unique names for each data set per project. To analyze the same experiment more than once, change the name of the experiment or change the **Project Name**.

f Select whether to **Use both Direct and Indirect relationships** for the analysis.

If you select **Yes**, IPA builds networks using both direct and indirect molecular interactions between genes. If you select **No**, IPA builds networks using only direct interactions between genes.

g Type in specific **Knowledge Base content**, if applicable.

Knowledge Base content indicates which database is searched for information to build the network. An empty string indicates to search all available Knowledge Bases and to incorporate information from all sources during the analysis.

h Select whether to **Include 'My Pathways' in Enrichment Score**.

If you select **Yes**, all pathways saved under My Pathways in IPA are included in the scoring process.

i Select whether to **Review Settings and ID Mapping before Running Analysis**.

If you select **Yes**, you can review and modify settings before running your IPA analysis. If you select **No**, IPA data analysis is automatically performed using the settings defined in this dialog box

j Select the **Gene Identifier Column**. The gene identifier is used to map genes in the entity list to genes in the IPKB.

k Click **OK**. Your default Internet browser is automatically launched and connected to the IPA server as specified in the **IPA Server Address**. See Figure 93 on page 79.

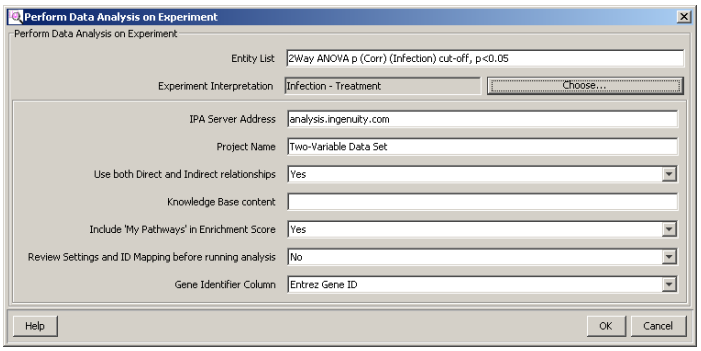


Figure 96 *Perform Data Analysis on Experiment dialog box*

5. Enter the options for **Perform Data Analysis on Entity List**.

a Review the **Entity List**. The active entity list is selected. To use a different entity list, cancel the operation, select a different entity list in the Experiment Navigator, and relaunch the operation.

b Click **Choose** to select the **Experiment Interpretation**. By default, the active interpretation is already selected (Figure 95).

Log_2 values for the conditions in the selected experiment interpretation are sent to IPA for analysis. The name of the data set used in IPA is named after the source experiment in Mass Profiler Professional.

c Click **OK**.

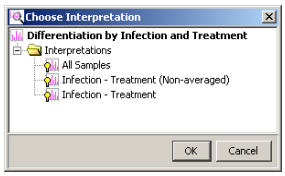


Figure 97 *Choose Interpretation dialog box*

d Review the **IPA Server Address**. Type in the address for the IPA server, for example, `analysis.ingenuity.com`.

e Type the name for the project to be created in **Project Name**. By default, the name used for the experiment that was originally selected is used. The **Project Name** is used by IPA under to store the pathway information.

Note: IPA only allows unique names for each data set per project. To analyze the same entity list more than once, change the name of the experiment or change the **Project Name**.

f Select whether to **Use both Direct and Indirect relationships** for the analysis.

If you select **Yes**, IPA builds networks using both direct and indirect molecular interactions between genes. If you select **No**, IPA builds networks using only direct interactions between genes.

g Type in specific **Knowledge Base content**, if applicable.

Knowledge Base content indicates which database is searched for information to build the network. An empty string indicates to search all available Knowledge Bases and to incorporate information from all sources during the analysis.

h Select whether to **Include 'My Pathways' in Enrichment Score**.

If you select **Yes**, all pathways saved under My Pathways in IPA are included in the scoring process.

i Select whether to **Review Settings and ID Mapping before Running Analysis**.

If you select **Yes**, you can review and modify settings before running your IPA analysis. If you select **No**, IPA data analysis is automatically performed using the settings defined in this dialog box

j Select the **Gene Identifier Column**. The gene identifier is used to map genes in the entity list to genes in the IPKB.

k Click **OK**. Your default Internet browser is automatically launched and connected to the IPA server as specified in the **IPA Server Address**. See Figure 93 on page 79.

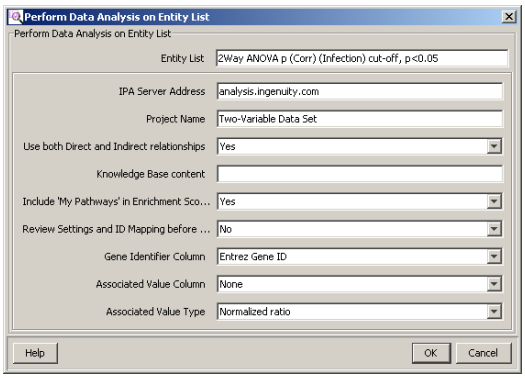


Figure 98 *Perform Data Analysis on Entity List dialog box*

Export to MetaCore

In order to use the **Export to MetaCore** operation your technology must contain Entrez Gene ID annotation. This operation is available for gene probe-based entity lists , not for compound-based entity lists .

Entrez is a cross-database search system that integrates the PubMed database of biomedical literature with other literature and molecular databases including DNA and protein sequence, structure, gene, genome, genetic variation, and gene expression. The Entrez search system is comprised of forty (40) molecular and literature databases and grows with advances in biomedical research. Entrez is maintained by the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/gquery).

Note: You must have an account with Thomson Reuters System Biology Solutions in order to make use of the Export to MetaCore operation. More information is available at Thomson Reuters Systems Biology (http://thomsonreuters.com/products_services/science/systems-biology/).

1. Launch **Export to MetaCore** in the Workflow Browser.

a Click **Export to MetaCore** in the Workflow Browser.

This operation is illustrated with "Agilent Expression Single Color Demo" sample data provided with your Mass Profiler Professional installation. The data is initially imported and analyzed following the "Creating an expression analysis using the sample array experiment" on page 23 of this workflow guide.

The Export to MetaCore operation has two (2) steps as shown in Figure 99 on page 83.

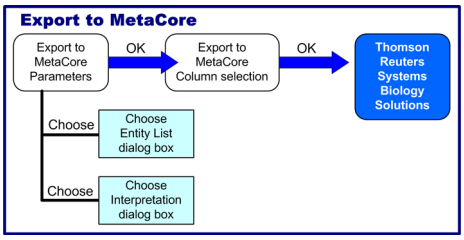


Figure 99 *Flow chart of the Export to MetaCore operation.*

b Click **OK** in the **Error** dialog box if you inadvertently launched Export to MetaCore when a compound-based entity list was the active entity list.

Figure 100 *Error dialog box*

2. Select and enter the parameters in the **Export to MetaCore** dialog box.

a Review the **Entity List**. The active entity list is selected.

b Click **Choose** to select a different **Entity List**. The entity list must be a probe-based entity list . You can select the **All Entities** entity list to send all the data in the experiment to MetaCore.

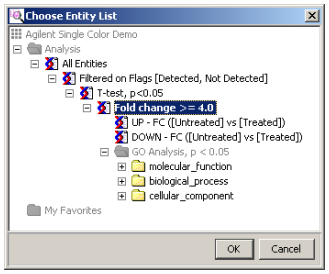


Figure 101 *Choose Entity List dialog box*

c Click **OK**.

d Review the **Interpretation**. The active interpretation list is selected.

e Click **Choose** to select a different **Interpretation**. The interpretation allows you to control which type of data is sent to MetaCore (sample-wise or condition-wise, average or non-averaged). If averaged data is selected the intensity values are averaged across samples in that condition. If a non-averaged interpretation is chosen, then you can send data one sample at a time.

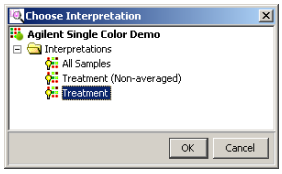


Figure 102 *Choose Interpretation dialog box*

f Click **OK**.

g Type the **MetaCore Server Address**. The default address is `http://portal.genego.com`. The address can be changed to point to your organization's installation. You must have a valid account on the server in order to be able to login to the portal at the end of this process.

h Type your **Experiment prefix**. The exported data is contained within an experiment in MetaCore and this option sets a prefix string to name the experiment. The default is a time-stamped string.

i Select the **Gene Identifier Column**. This option sets the identifier of the data column that is exported to MetaCore. Currently, there is only one option: **Entrez Gene ID**.

Note: If the Entrez Gene ID annotation is not present for the technology of the chosen entity list you must update the technology with Entrez Gene ID annotations before proceeding.

j Click **OK**.

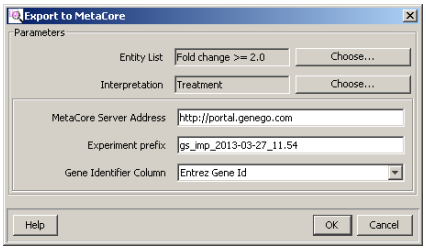


Figure 103 *Export to MetaCore dialog box, Parameters*

3. Enter the column selection in the **Export to MetaCore** dialog box.

a Select the data column for the **Choose data column**.

b Type in an **Experiment suffix to be added**. The column name is used if no characters are entered.

c Click **OK**.

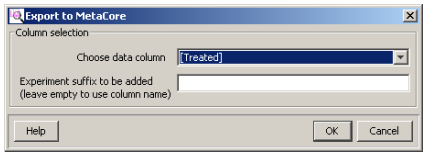


Figure 104 *Export to MetaCore dialog box, Column selection*

4. Approve opening a browser window to log into MetaCore.

a Click **OK**. Your default browser is launched.

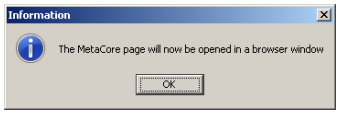


Figure 105 *Information dialog box indicating that a browser window is required*

5. Log into MetaCore.

a Review the progress of your browser window. A submittal notice is displayed by your browser as shown in Figure 106 before you are directed to the MetaCore site.

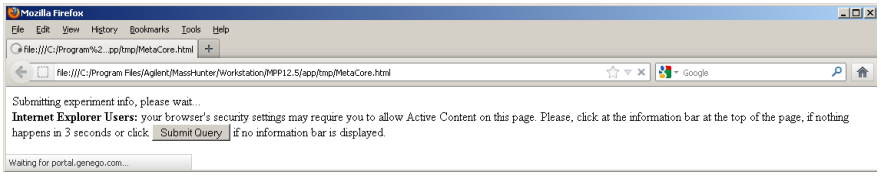

Mozilla Firefox
Submitting experiment info, please wait...
Internet Explorer Users: your browser's security settings may require you to allow Active Content on this page. Please, click at the information bar at the top of the page, if nothing happens in 3 seconds or click Submit Query if no information bar is displayed.
Waiting for portal.genego.com...

Figure 106 *MetaCore submittal notice in your browser*

b Enter your **Username** and **Password** to log into your MetaCore account.

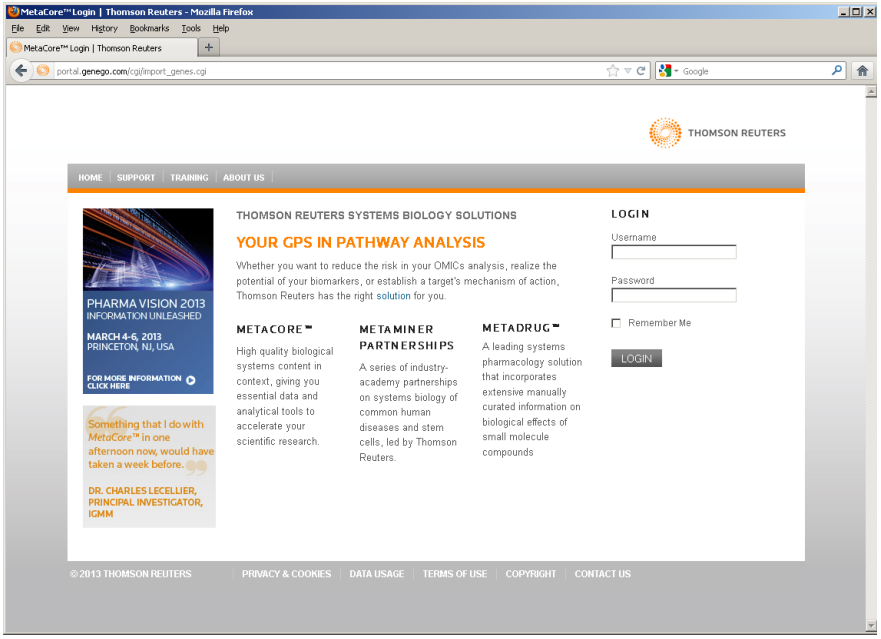

MetaCore™ Login | Thomson Reuters - Mozilla Firefox
THOMSON REUTERS
HOME SUPPORT TRAINING ABOUT US
THOMSON REUTERS SYSTEMS BIOLOGY SOLUTIONS
YOUR GPS IN PATHWAY ANALYSIS
Whether you want to reduce the risk in your OMICs analysis, realize the potential of your biomarkers, or establish a target's mechanism of action, Thomson Reuters has the right solution for you.
METACORE™
High quality biological systems content in context, giving you essential data and analytical tools to accelerate your scientific research.
METAMINER PARTNERSHIPS
A series of industry-academy partnerships on systems biology of common human diseases and stem cells, led by Thomson Reuters.
METADRUG™
A leading systems pharmacology solution that incorporates extensive manually curated information on biological effects of small molecule compounds
LOGIN
Username
Password
Remember Me
LOGIN
PHARMA VISION 2013
INFORMATION UNLEASHED
MARCH 4-6, 2013
PRINCETON, NJ, USA
FOR MORE INFORMATION CLICK HERE
Something that I do with MetaCore™ in one afternoon now, would have taken a week before.
DR. CHARLES LECELLIER, PRINCIPAL INVESTIGATOR, IGMM
© 2013 THOMSON REUTERS
PRIVACY & COOKIES DATA USAGE TERMS OF USE COPYRIGHT CONTACT US

Figure 107 *MetaCore login page*

c The export process is now complete.

Connect to Cytoscape

Cytoscape is a biological network visualization and analysis tool. Cytoscape is used to visualize molecular interaction networks and provide you with a means to generate views of gene and protein associations. Cytoscape is built on an open source platform and no cost to download and use with Mass Profiler Professional. The Agilent Cytoscape plug-in files enable the feature to send entity lists from your active MPP experiment to Cytoscape.

Note: The Connect to Cytoscape features in Mass Profiler Professional are part of GeneSpring GX. If your GeneSpring GX module license does not include **Connect to Cytoscape** contact Agilent support (click **Help > Contact Technical Support** on the menu bar) for assistance. **Connect to Cytoscape** is a separate feature and can only be accessed with a valid **GeneSpring GX** module license. See "Getting started requirements" on page 62

The Connect to Cytoscape operation does not have an intermediate wizard or dialog box like the other Pathway Analysis operations. If Connect to Cytoscape is an active feature in your installation, Mass Profiler Professional immediately starts transferring the entity list from your active experiment and launches Cytoscape when the operation is invoked. It is recommended to review all of the steps in this operation before selecting the Connect to Cytoscape operation to make sure your installation of MPP and Cytoscape are enabled to work together.

---

1. Determine if **Connect to Cytoscape** is an active feature of MPP.

Connect to Cytoscape is a separate feature and can only be accessed with a valid **GeneSpring GX** module license. This step determines if your GeneSpring GX license includes **Connect to Cytoscape**.

a Click **Tools > Options** on the menu bar to launch configuration options.

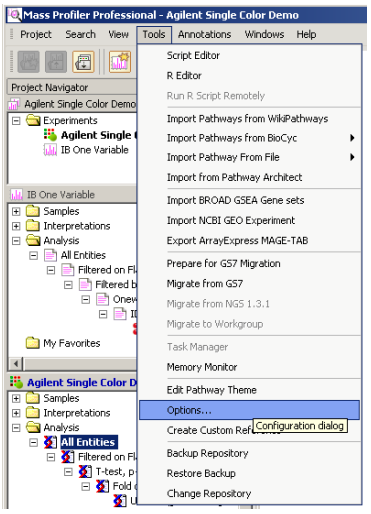


Figure 108 *Launching the configuration options from the menu bar*

b Click **Miscellaneous** on the left-hand pane in the **Configuration Dialog** dialog box.

c Click **Cytoscape Installation Path** on the left-hand pane in the **Configuration Dialog** dialog box.

d Determine if you have a user entry field to type a **Cytoscape installation path** (see Figure 109 on page 87). If the entry field is available continue to the next step.

Note: If the entry field **Cytoscape installation path** not available, stop at this step and contact Agilent support to activate this feature.

---

2. Enter the **Cytoscape Installation Path**.

a Click **Browse** to select the **Cytoscape installation path** in the **Configuration Dialog**.

b Select the folder that contains the Cytoscape program (Cytoscape.exe) in the **Choose a File** dialog box (see Figure 61 on page 57 for a typical dialog box).

c Click **Open**.

d Click **OK**.

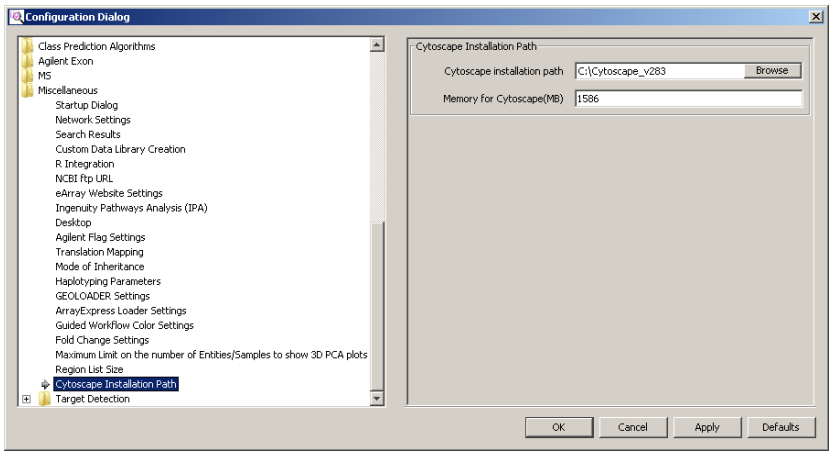


Figure 109 *Cytoscape Installation Path in the Configuration Dialog*

3. Launch **Connect to Cytoscape** in the Workflow Browser.

When passing entities to Cytoscape, the active entity list must belong to the active experiment. An overview of the Project Navigator and Experiment Navigator functional areas within Mass Profiler Professional are shown in Figure 29 on page 37; the active entity list and experiment are in bold font.

a Click **Connect to Cytoscape** in the Workflow Browser. MPP immediately starts transferring your entity list and launches Cytoscape.

b Click **Cancel** in the **Send Entities to Cytoscape** progress box if you launched Connect to Cytoscape with an active entity list that does not belong to the active experiment, or if you want to stop **Connect to Cytoscape** for another reason.

Note: If the **Send Entities to Cytoscape** progress box indicates that "GeneSpring was unable to start Cytoscape" or a message indicating that "Access is denied," stop and continue to step 5 "Download Cytoscape 2.8.x to your computer." on page 88 and then return to this step.

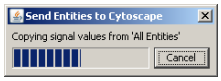


Figure 110 *The normal Send Entities to Cytoscape progress box.*

4. Perform your analysis with **Cytoscape**.

When Cytoscape is launched a splash screen is displayed with the version number as shown in Figure 111 on page 87.

Figure 111 *Cytoscape splash screen at startup*

a Begin using Cytoscape to perform your network visualization and analysis. Cytoscape is used to visualize molecular interaction networks and provide you with a means to generate views of gene and protein associations.

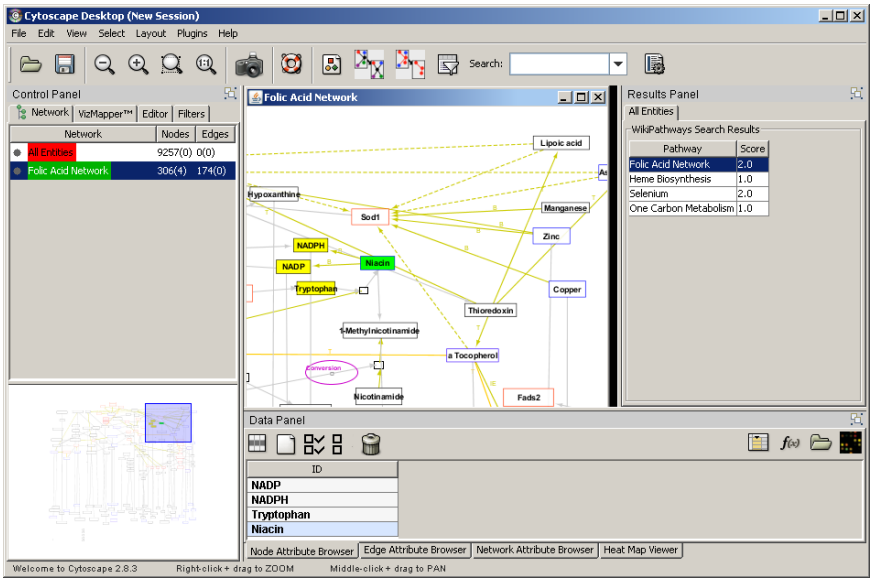


Figure 112 *Cytoscape Desktop*

b Click **Help > Contents**, or press **F1**, to access the Cytoscape User Manual for information on how to use Cytoscape (Figure 113).

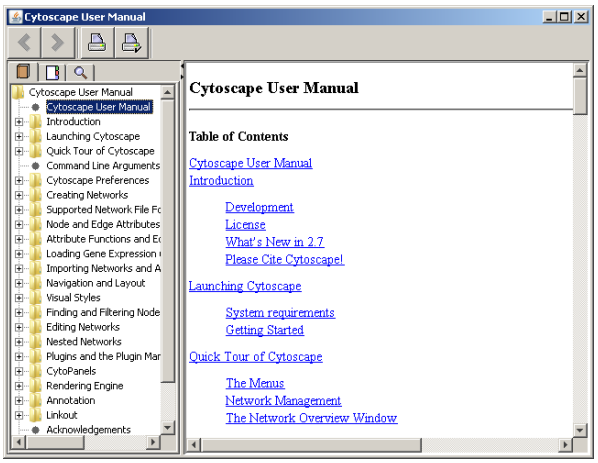


Figure 113 *Cytoscape User Manual accessed from* ***Help > Contents***

c The connection process is now complete. You can continue analyzing your data with Mass Profiler Professional at the same time your Cytoscape session is running.

d Close Cytoscape before starting a new analysis. Re-launching **Connect to Cytoscape** while Cytoscape is still open with a prior entity list adds the new entity list to the prior analysis and experiment within Cytoscape.

---

5. Download **Cytoscape 2.8.x** to your computer.

There is no cost to register, download, and install Cytoscape on your computer.

a Close Mass Profiler Professional.

b Open http://www.cytoscape.org in your Internet browser.

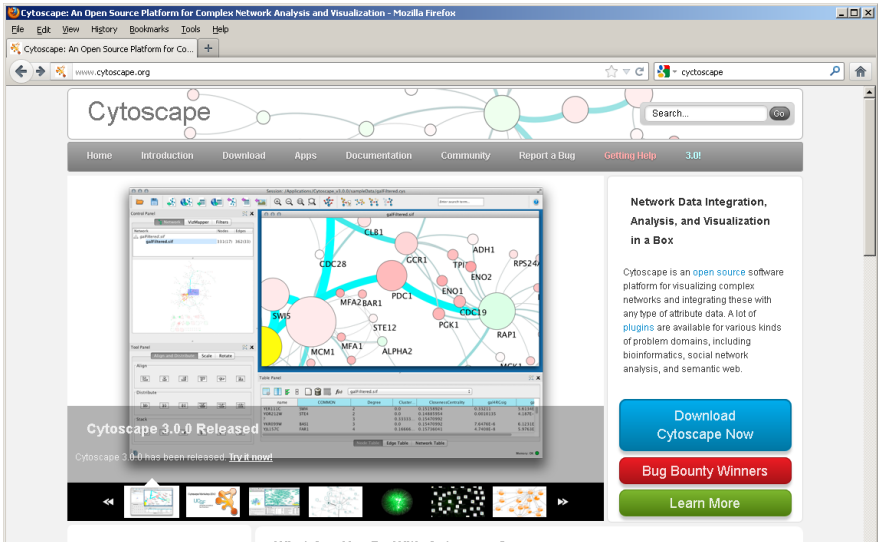

Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization - Mozilla Firefox
Cytoscape
Home
Introduction
Download
Apps
Documentation
Community
Report a Bug
Getting Help
Network Data Integration, Analysis, and Visualization in a Box
Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute data. A lot of plugins are available for various kinds of problem domains, including bioinformatics, social network analysis, and semantic web.
Download Cytoscape Now
Bug Bounty Winners
Learn More
Cytoscape 3.0.0 Released

Figure 114 *Cytoscape web site*

c Click **Download Cytoscape Now**.

d Type in your information and accept the terms of the License Agreement on the Cytoscape download page.

e Click **Proceed to Download**.

f Download the Latest Product Version 2.8.x (version 2.8.1 or higher) and install it in a directory that has all read and write permissions available.

Note: Cytoscape version 3.x may not be compatible with Connect to Cytoscape. Contact Agilent support to see if 3.x is supported.

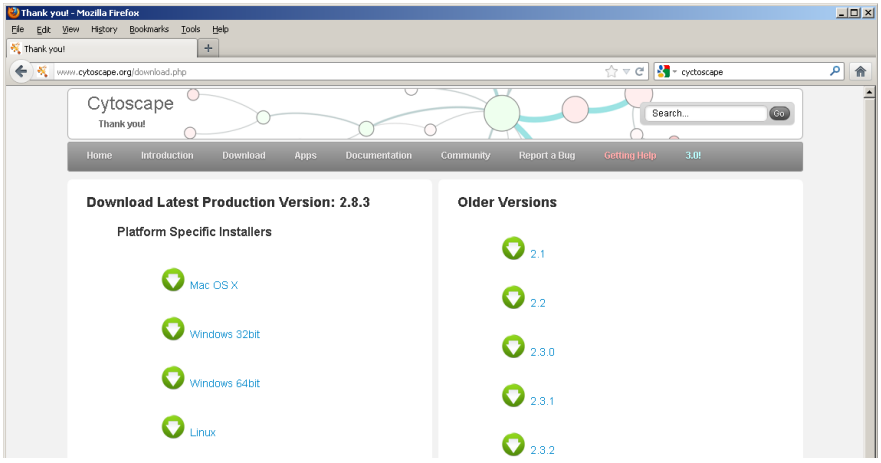


Figure 115 *Cytoscape download web site*

6. Install **Cytoscape** on your computer.

a Run the installation file downloaded during the prior step.

Note: The Cytoscape installation directory path cannot have any spaces. Choose or create a new directory path from "C:\" to install Cytoscape. Do not install Cyto-

scape in the default "C:\Program Files" directory since there is a space in this directory path.

b Download and install the Java Runtime Environment, if you are prompted.

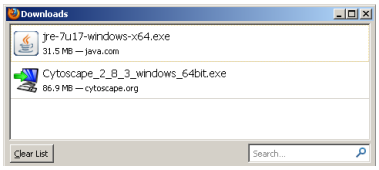


Figure 116 *Downloaded Cytoscape installation file and Java Runtime Environment installation file.*

---

7. Configure **Cytoscape** to work with Mass Profiler Professional.

In order to enable Mass Profiler Professional to transfer data and launch Cytoscape you must download the *Cytoscape_Patch_n_Plugins.zip* file and follow the included instructions.

a Open http://basil.strandls.com/downloads/Cyto/ in your browser.

b Click **Cytoscape_Patch_n_Plugins.zip**.

c Select **Save File** in the **Opening Cytoscape_Patch_n_Plugins.zip** dialog box.

d Click **OK**. *Cytoscape_Patch_n_Plugins.zip* is saved to your downloads folder on your PC.

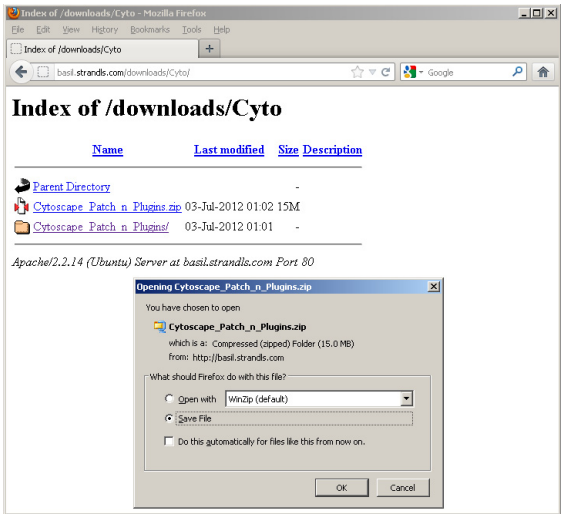

Index of /downloads/Cyto - Mozilla Firefox
Index of /downloads/Cyto
basil.strandls.com/downloads/Cyto/
Index of /downloads/Cyto
Name Last modified Size Description
Parent Directory -
Cytoscape_Patch_n_Plugins.zip 03-Jul-2012 01:02 15M
Cytoscape_Patch_n_Plugins/ 03-Jul-2012 01:01 -
Apache/2.2.14 (Ubuntu) Server at basil.strandls.com Port 80
Opening Cytoscape_Patch_n_Plugins.zip
You have chosen to open
Cytoscape_Patch_n_Plugins.zip
which is a: Compressed (zipped) Folder (15.0 MB)
from: http://basil.strandls.com
What should Firefox do with this file?
Open with WinZip (default)
Save File
Do this automatically for files like this from now on.
OK
Cancel

Figure 117 *Cytoscape_Patch_n_Plugins.zip file location and Save File*

e Open *Cytoscape_Patch_n_Plugins.zip*. The files included in the zip file are:

AdaptiveJavaHelp.jar
CriteriaMapper.jar
CytoscapeConnector-1.0-SNAPSHOT.jar
GeneSpringConnector-1.0-SNAPSHOT.jar
GOElitePlugin.jar
gpml.jar
HeatMapViewer-2.2.1.jar
HeatStripPlugin.jar
PathwaySearchPluginWithLibs.jar

SendGenesAndEnrichmentFilesToCytoscape$py.class
SendMetabolitesAndInterpretationToCytoscape$py.class
README.txt

f Copy the nine (9) jar files to the *\plugins* folder in your Cytoscape installation directory.

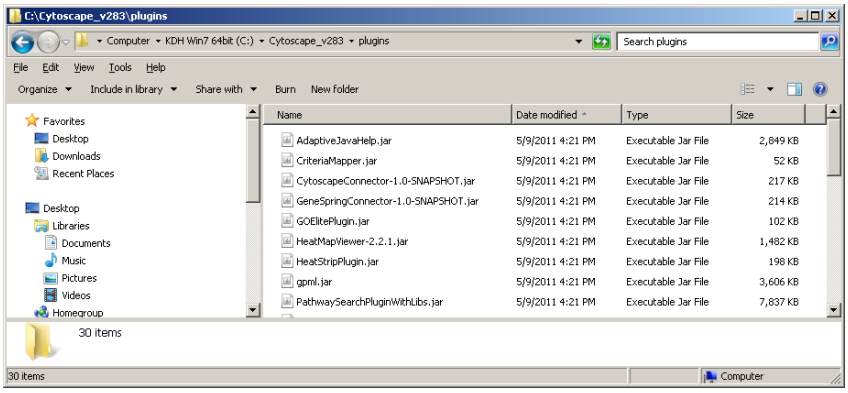

C:\Cytoscape_v283\plugins
Computer KDH Win7 64bit (C:) Cytoscape_v283 plugins
Search plugins
File Edit View Tools Help
Organize Include in library Share with Burn New folder
Favorites
Desktop
Downloads
Recent Places
Desktop
Libraries
Documents
Music
Pictures
Videos
Homegroup
Name Date modified Type Size
AdaptiveJavaHelp.jar 5/9/2011 4:21 PM Executable Jar File 2,849 KB
CriteriaMapper.jar 5/9/2011 4:21 PM Executable Jar File 52 KB
CytoscapeConnector-1.0-SNAPSHOT.jar 5/9/2011 4:21 PM Executable Jar File 217 KB
GeneSpringConnector-1.0-SNAPSHOT.jar 5/9/2011 4:21 PM Executable Jar File 214 KB
GOEllitePlugin.jar 5/9/2011 4:21 PM Executable Jar File 102 KB
HeatMapViewer-2.2.1.jar 5/9/2011 4:21 PM Executable Jar File 1,482 KB
HeatStripPlugin.jar 5/9/2011 4:21 PM Executable Jar File 198 KB
gpml.jar 5/9/2011 4:21 PM Executable Jar File 3,606 KB
PathwaySearchPluginWithLibs.jar 5/9/2011 4:21 PM Executable Jar File 7,837 KB
30 items
30 items
Computer

Figure 118 *Jar files copied to the Cytoscape \plugins folder*

g Copy the two (2) class files to the *\bin\packages\marray\cytoscape\1.0\scripts* folder in your MPP installation directory.

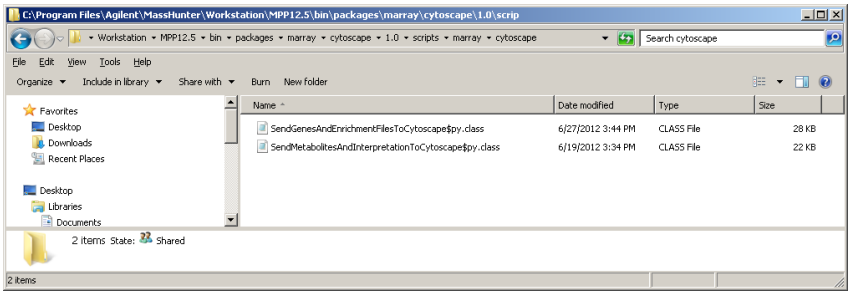
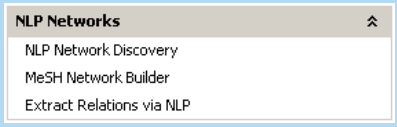


Figure 119 *Class files copied to the MPP \bin\packages\marray\cytoscape\1.0\scripts folder*

h Run Mass Profiler Professional and open your recent project.

i Go to step 2"Enter the Cytoscape Installation Path." on page 86 to configure Cytoscape and then launch **Connect to Cytoscape**.

NLP Networks



NLP Networks drives discovery by creating networks around the entities of interest using a powerful Natural Language Processing (NLP) algorithm that extracts information from published literature. The operations available help you to create pathways from PubMed abstracts, the MeSH (Medical Subject Headings) database, selected entities, or personal data sources using NLP

Note: The NLP Networks features in Mass Profiler Professional are part of the Pathway Analysis module. Pathway Analysis is licensed separately and can only be accessed with a valid Pathway Architect module license. See "Getting started requirements" on page 62.

NLP Networks consists of three operations:

- "NLP Network Discovery" on page 93
- "MeSH Network Builder" on page 99
- "Extract Relations via NLP" on page 102

Useful supplemental task also documented:

- "Create Pathway Organism" on page 106

NLP Networks features

Create networks based on information in PubMed abstracts and identify interactions associated with Medical Subject Headings (MeSH) terms using NLP as an alternate way of creating pathways based on terms and concepts instead of entities. Once you have created and saved such networks in Mass Profiler Professional you can overlay data from your experiments on these networks to help you find significant pathways and networks.

NLP uses a method that carefully parses your sentence structure to maximize accuracy and control different aspects of a sentence without compromising recall reliability. The NLP system operates on a sentence-by-sentence manner and extracts only those relations that are completely within a sentence. NLP employs four main phases to ensure accuracy - entity recognition, syntax analysis, semantic analysis, and semantic inference. See section "12.1 Natural Language Processing (NLP) in Mass Profiler Professional" in the *Mass Profiler Professional User Manual* for more information.

Interaction Databases

The Pathway Analysis module is integrated with a database of relations between various biological molecules and processes. The molecules and processes are depicted as **Entities** and their biological interactions as **Relations**. In a pathway view entities form the nodes of the graph and the lines depict the relations. An organism entity database consists of proteins, small molecules, processes, functions, enzymes, complexes, and families. Proteins are organism specific while the other entities of the organism are largely organism independent.

The Interaction Database is organized in a hierarchical fashion with two levels. The top level is generic and contains information that is common across organisms. The second level comprises the various organism specific entities (predominantly proteins) and relations specific to the organism. The public sources used by the interaction databases is described in section "12.2.2 Database Entities" in the *Mass Profiler Professional User Manual*.

You can download and update Interaction Databases from the Agilent Server or with a Mass Profiler Professional update file. If you are working with an organism that is not currently available in Mass Profiler Professional you can create a new organism; click **Annotations > Create Pathway Organism**. Valid taxonomy IDs can be found at the Entrez Taxonomy database site (http://www.ncbi.nlm.nih.gov/taxonomy). See "Create Pathway Organism" on page 106 to add a new organism.

NLP Network Discovery

NLP Network Discovery is performed on entity lists and selected entities in a pathway viewer. To perform NLP Network Discovery on custom lists of entities you can create a Pathway experiment.

The queried database corresponds to the organism specified in the technology of the current experiment. Mass Profiler Professional uses Entrez Gene ID, Swiss-Prot, and Gene Symbol from the technology for this query to map to available Entrez Gene IDs, and available entries in the Protein field, and the Symbol field of the Interaction Database, respectively. It is important that both the technology and the Interaction Database contain at least one of these annotations.

The NLP Network Discovery operation has two options for exploring the most common functionalities of network discovery:

Simple Analysis: Provides you with a selection of the most common functionalities of network discovery. The default settings for guiding you through the simple network discovery workflow include:

- matching the selected entities to entities in the database
- retrieving relevant relations between the set of matched entities
- displaying the results in a network graphical view

Advanced Analysis: Enables you to change and specify the details at every step of the network discovery process.

Organism specific Interaction Databases are available as updates to Mass Profiler Professional. The relations in the database are mainly derived from published literature abstracts using NLP.

1. Launch **NLP Network Discovery** in the Workflow Browser.

a Click **NLP Network Discovery** in the Workflow Browser.

This operation is illustrated with "Agilent Expression Single Color Demo" sample data provided with your Mass Profiler Professional installation. The data is initially imported and analyzed following the "Creating an expression analysis using the sample array experiment" on page 23 of this workflow guide.

The NLP Network Discovery operation has five (5) steps as shown in Figure 120 on page 94. The steps that you use depend on your selection for the **Analysis Type** in the first step of the wizard. The new pathway list is placed in the Analysis folder within the Experiment Navigator.

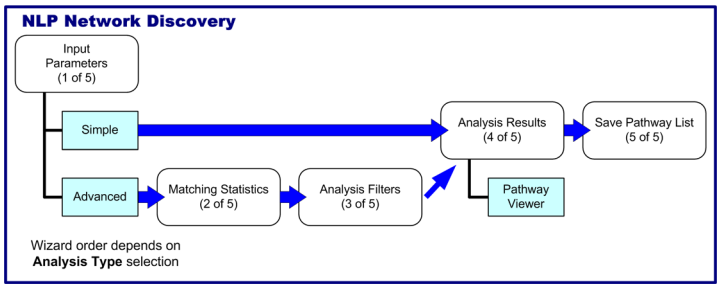

NLP Network Discovery
Input Parameters (1 of 5)
Simple
Advanced
Matching Statistics (2 of 5)
Analysis Filters (3 of 5)
Analysis Results (4 of 5)
Pathway Viewer
Save Pathway List (5 of 5)
Wizard order depends on Analysis Type selection

Figure 120 *Flow chart of the NLP Network Discovery operation.*

2. Input parameters in **NLP Network Discovery (Step 1 of 5)**.

a Select an **Input List**. By default the active entity list is selected. The active entity list must belong to the active experiment otherwise an error indicating "No relations found" may be encountered.

b Select an **Analysis Type** from the two choices. Your selection determines the available option for **Algorithm** and the steps through the wizard as shown in Figure 120.

Simple Analysis: Provides you with a selection of the most common functionalities of network discovery. The default settings for guiding you through the simple network discovery workflow include:

Advanced Analysis: Enables you to change and specify the details at every step of the network discovery and pathway creation process.

c Select an **Algorithm**.

1. If you selected **Simple** for the **Analysis Type** your options are:

Direct Interactions: Find relations that connect the selected entities.

Network Targets and Regulators: Find entities that are upstream and downstream of two or more entities from the original list.

Network Targets: Find downstream entity targets that connect two or more entities from the original list of selected entities.

Network Regulators: Find upstream entity regulators that connect to two or more entities from the original list of selected entities.

Network Binders: Find entities that "bind" (entities that are connected by binding interactions) to two or more entities from the original list of selected entities.

Network Modifiers: Find protein entities that are either regulators or targets of biochemical protein modifications of two or more proteins from the original list of selected entities.

Transcription Regulators: Find protein entities regulating mRNA expression of, or whose expression is regulated by two or more entities from the original list of selected entities.

Transport Regulators: Find all compounds that are regulating the transport of other compounds.

Metabolism Regulators: Find compounds that are regulating the metabolism of biomolecules.

Small Molecules: Find all small molecules (drugs) regulators and targets of two or more entities from the original list of selected entities.

Biological Processes: Finds all biological process entities connected to two or more entities from the original list of selected entities.

Shortest Connect: Finds the smallest set of relations that connects all entities in a given list into a single network.

In addition to the Algorithm selection above, **Simple** also sets the following parameters:

Algorithms type: local/global
Connectivity: Connectivity relevance 50 and Connectivity <= 2
Entity filter: All entities are selected
Quality filter: >=9
Relation filter: All relation types are selected

2. If you selected **Advanced** for the **Analysis Type** your options are:

Direct Interactions: Find relations that connect the selected entities.

Expand Interactions: Expand the existing network to include the first-degree neighbors of the selected entities.

Shortest Connect: Find the smallest set of relations that connects the set of selected entities into a single network. Some intermediate entities may be introduced in this process.

d Click **Next**.

Go to step 5 "Review the analysis results in NLP Network Discovery (Step 4 of 5)." on page 98 if your **Analysis Type** is **Simple**.

Go to step 3 "Select matching entities in NLP Network Discovery (Step 2 of 5)." if your **Analysis Type** is **Advanced**.

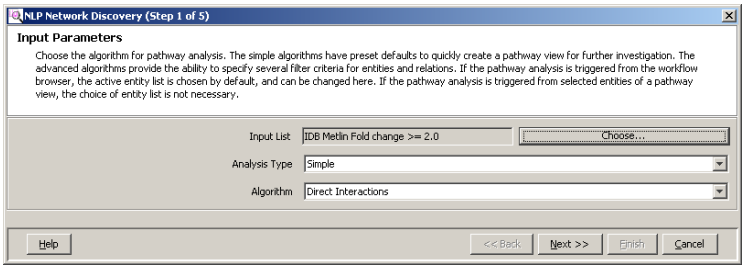

NLP Network Discovery (Step 1 of 5)
Input Parameters
Choose the algorithm for pathway analysis. The simple algorithms have preset defaults to quickly create a pathway view for further investigation. The advanced algorithms provide the ability to specify several filter criteria for entities and relations. If the pathway analysis is triggered from the workflow browser, the active entity list is chosen by default, and can be changed here. If the pathway analysis is triggered from selected entities of a pathway view, the choice of entity list is not necessary.
Input List
IDB Metlin Fold change >= 2.0
Choose...
Analysis Type
Simple
Algorithm
Direct Interactions
Help
<< Back
Next >>
Finish
Cancel

Figure 121 *Input Parameters page (NLP Network Discovery (Step 1 of 5))*

3. Select matching entities in **NLP Network Discovery (Step 2 of 5)**.

This step is only encountered if you selected **Advanced** for the **Analysis Type** in "Input parameters in NLP Network Discovery (Step 1 of 5)." If the algorithm you selected does not find any entities that meet the algorithm criteria you are prompted to select a different algorithm or another input list for analysis.

a Review the matched, not matched, and redundant entities and their related statistics.

b Select any or all of the entities to use in your pathway analysis. When an entity list is selected the row is highlighted.

Select a continuous range of entity lists - click on the first file and press **Shift** and click on the last entity list that includes the range of entity lists you want to select.

Select discontinuous or individual entity list - press **Ctrl** and click on additional entity lists.

c Click **Next**.

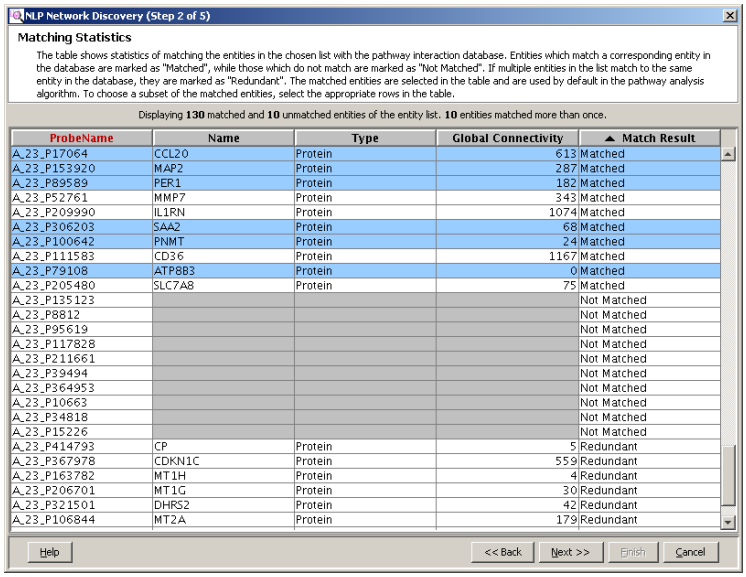

NLP Network Discovery (Step 2 of 5)
Matching Statistics
The table shows statistics of matching the entities in the chosen list with the pathway interaction database. Entities which match a corresponding entity in the database are marked as "Matched", while those which do not match are marked as "Not Matched". If multiple entities in the list match to the same entity in the database, they are marked as "Redundant". The matched entities are selected in the table and are used by default in the pathway analysis algorithm. To choose a subset of the matched entities, select the appropriate rows in the table.
Displaying 130 matched and 10 unmatched entities of the entity list. 10 entities matched more than once.
ProbeName | Name | Type | Global Connectivity | Match Result
A_23_P17064 | CCL20 | Protein | 613 | Matched
A_23_P153920 | MAP2 | Protein | 287 | Matched
A_23_P89589 | PER1 | Protein | 182 | Matched
A_23_P52761 | MMP7 | Protein | 343 | Matched
A_23_P209990 | IL1RN | Protein | 1074 | Matched
A_23_P306203 | SAA2 | Protein | 68 | Matched
A_23_P100642 | PNMT | Protein | 24 | Matched
A_23_P111583 | CD36 | Protein | 1167 | Matched
A_23_P79108 | ATP8B3 | Protein | 0 | Matched
A_23_P205480 | SLC7A8 | Protein | 75 | Matched
A_23_P135123 | | | | Not Matched
A_23_P8812 | | | | Not Matched
A_23_P95619 | | | | Not Matched
A_23_P117828 | | | | Not Matched
A_23_P211661 | | | | Not Matched
A_23_P39494 | | | | Not Matched
A_23_P364953 | | | | Not Matched
A_23_P10663 | | | | Not Matched
A_23_P34818 | | | | Not Matched
A_23_P15226 | | | | Not Matched
A_23_P414793 | CP | Protein | 5 | Redundant
A_23_P367978 | CDKN1C | Protein | 559 | Redundant
A_23_P163782 | MT1H | Protein | 4 | Redundant
A_23_P206701 | MT1G | Protein | 30 | Redundant
A_23_P321501 | DHRS2 | Protein | 42 | Redundant
A_23_P106844 | MT2A | Protein | 179 | Redundant
Help
<< Back
Next >>
Finish
Cancel

Figure 122 *Matching Statistics page (NLP Network Discovery (Step 2 of 5))*

4. Select and enter filter parameters in **NLP Network Discovery (Step 3 of 5)**.

This step is only encountered if you selected **Advanced** for the **Analysis Type** in "Input parameters in NLP Network Discovery (Step 1 of 5)."

a Type in the **Relation score** for your analysis. The score is a value between 1 and 10 with 10 indicating the highest score, the best quality. The default value is >= 9.

b Mark the relations to include in your analysis in **Select relation type**.

If your **Algorithm** is **Expanded Interactions** (see Figure 124 on page 97):

c Type in the **Entity local connectivity** for your analysis. This is a filter that specifies the number of entities in the input list that a new entity must be connected to in oder for the new entity to be added to the list. The default value is >= 2.

d Mark the types of entities to evaluate in **Select entity type**.

e Select the **Limit analysis results based on**.

Local connectivity: Allows you to add a certain number of entities to the given network based on their rank with regards to local connectivity. New entities are ranked with decreasing priority, based on how many entities they are connected with within a given list of entities.

Local to global connectivity ratio: A local/global connectivity ratio is computed for each new entity. Local connectivity is based on the number of entities to which it connects within a given list and global connectivity is the number of relations that it participates in within the entire database. New entities are ranked

with decreasing priority based on this local/global connectivity ratio. This is the default value.

f Type in the **Maximum number of new entities** to limit the number of entities to add to your network. The default value is 50.

If your **Algorithm** is **Shortest Connect** (see Figure 125 on page 98):

g Type in the **Entity global connectivity** for your analysis. This is a filter that adds new entities to connect two disconnected network clusters based on the number global entities that must be connected to input entity list in oder for the new entity to be added to the list. The default value is <= 500.

h Mark the types of entities to evaluate in **Select entity type**.

i Click **Next**.

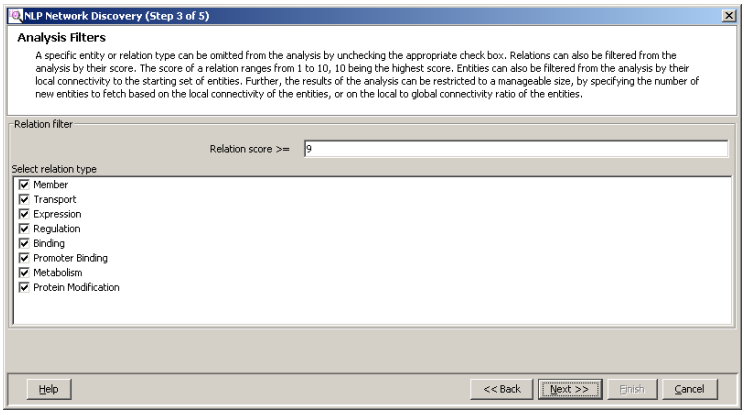

NLP Network Discovery (Step 3 of 5)
Analysis Filters
A specific entity or relation type can be omitted from the analysis by unchecking the appropriate check box. Relations can also be filtered from the analysis by their score. The score of a relation ranges from 1 to 10, 10 being the highest score. Entities can also be filtered from the analysis by their local connectivity to the starting set of entities. Further, the results of the analysis can be restricted to a manageable size, by specifying the number of new entities to fetch based on the local connectivity of the entities, or on the local to global connectivity ratio of the entities.
Relation filter
Relation score >=
9
Select relation type
Member
Transport
Expression
Regulation
Binding
Promoter Binding
Metabolism
Protein Modification
Help
<< Back
Next >>
Finish
Cancel

Figure 123 *Direct Interactions Analysis Filters page (NLP Network Discovery (Step 3 of 5))*

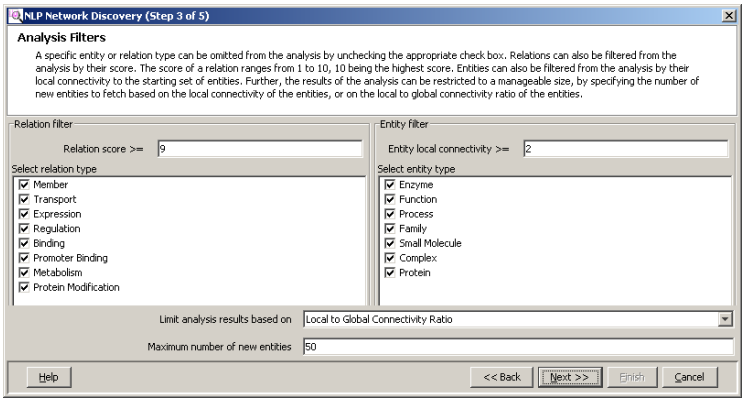

NLP Network Discovery (Step 3 of 5)
Analysis Filters
A specific entity or relation type can be omitted from the analysis by unchecking the appropriate check box. Relations can also be filtered from the analysis by their score. The score of a relation ranges from 1 to 10, 10 being the highest score. Entities can also be filtered from the analysis by their local connectivity to the starting set of entities. Further, the results of the analysis can be restricted to a manageable size, by specifying the number of new entities to fetch based on the local connectivity of the entities, or on the local to global connectivity ratio of the entities.
Relation filter
Relation score >=
9
Select relation type
Member
Transport
Expression
Regulation
Binding
Promoter Binding
Metabolism
Protein Modification
Entity filter
Entity local connectivity >=
2
Select entity type
Enzyme
Function
Process
Family
Small Molecule
Complex
Protein
Limit analysis results based on
Local to Global Connectivity Ratio
Maximum number of new entities
50
Help
<< Back
Next >>
Finish
Cancel

Figure 124 *Expanded Interactions Analysis Filters page (NLP Network Discovery (Step 3 of 5))*

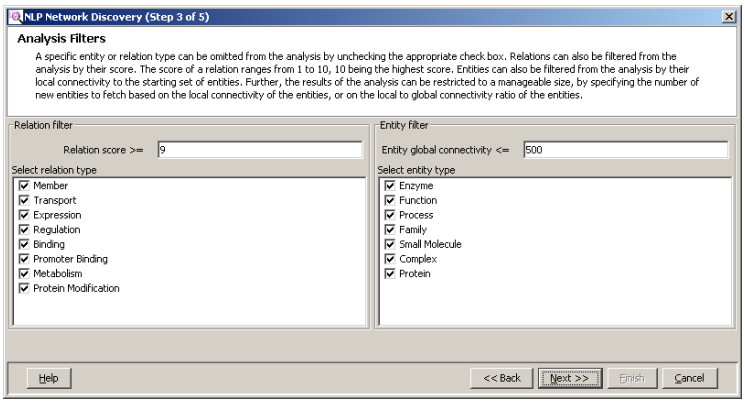

NLP Network Discovery (Step 3 of 5)
Analysis Filters
A specific entity or relation type can be omitted from the analysis by unchecking the appropriate check box. Relations can also be filtered from the analysis by their score. The score of a relation ranges from 1 to 10, 10 being the highest score. Entities can also be filtered from the analysis by their local connectivity to the starting set of entities. Further, the results of the analysis can be restricted to a manageable size, by specifying the number of new entities to fetch based on the local connectivity of the entities, or on the local to global connectivity ratio of the entities.
Relation filter
Relation score >= 9
Select relation type
Member
Transport
Expression
Regulation
Binding
Promoter Binding
Metabolism
Protein Modification
Entity filter
Entity global connectivity <= 500
Select entity type
Enzyme
Function
Process
Family
Small Molecule
Complex
Protein
Help
<< Back
Next >>
Finish
Cancel

Figure 125 *Shortest Connect Analysis Filters page (NLP Network Discovery (Step 3 of 5))*

5. Review the analysis results in **NLP Network Discovery (Step 4 of 5)**.

Analysis Result displays the created pathway. The initial number of entities, the number of new relations, and the number of new entities are displayed.

a Review the pathway.

b Edit the pathway. Details for using the pathway view is described in section "11.1.3 Creating and Editing Pathways" in the *Mass Profiler Professional User Manual*.

c Click **Next**.

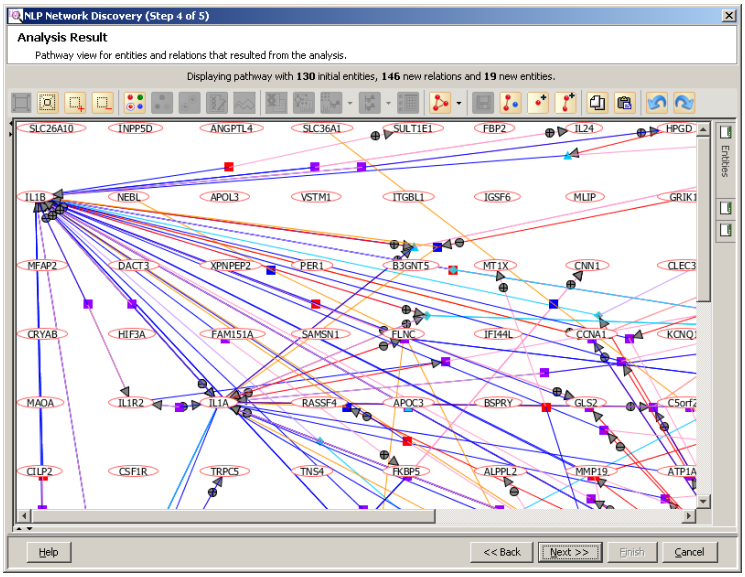


Figure 126 *Analysis Results page (NLP Network Discovery (Step 4 of 5))*

6. Save the pathway list in **NLP Network Discovery (Step 5 of 5)**.

Analysis Result displays the created pathway. The initial number of entities, the number of new relations, and the number of new entities are displayed.

a Review the pathway list.

b Type a descriptive **Name** that is stored with the saved pathway entity list.

c Edit the **Notes** that are stored with the saved pathway entity list.

d Double-click a row in the **Pathways** table to launch the Pathway Inspector to review the entities and relations contained in the new pathway.

e Click **Finish**.

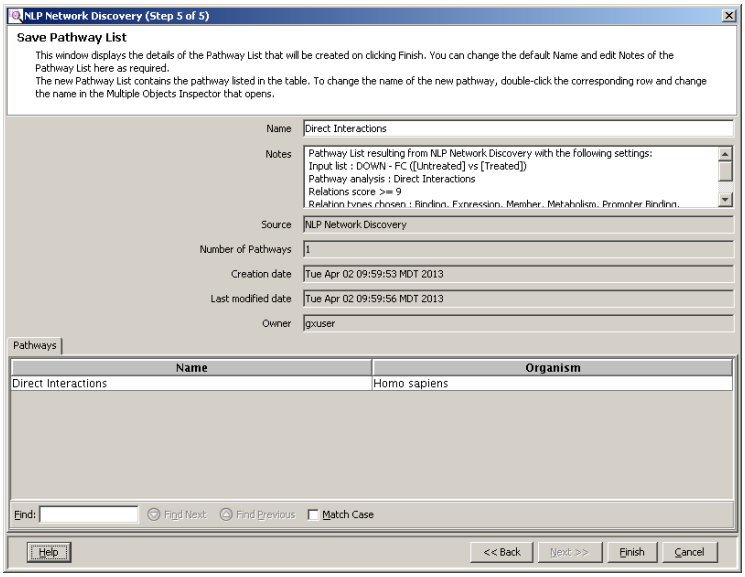

NLP Network Discovery (Step 5 of 5)
Save Pathway List
This window displays the details of the Pathway List that will be created on clicking Finish. You can change the default Name and edit Notes of the Pathway List here as required.
The new Pathway List contains the pathway listed in the table. To change the name of the new pathway, double-click the corresponding row and change the name in the Multiple Objects Inspector that opens.
Name
Direct Interactions
Notes
Pathway List resulting from NLP Network Discovery with the following settings:
Input list : DOWN - FC ([Untreated] vs [Treated])
Pathway analysis : Direct Interactions
Relations score >= 9
Source
NLP Network Discovery
Number of Pathways
1
Creation date
Tue Apr 02 09:59:53 MDT 2013
Last modified date
Tue Apr 02 09:59:56 MDT 2013
Owner
gxuser
Pathways
Name
Organism
Direct Interactions
Homo sapiens
Find:
Find Next
Find Previous
Match Case
Help
<< Back
Next >>
Finish
Cancel

Figure 127 *Analysis Results page (NLP Network Discovery (Step 5 of 5))*

MeSH Network Builder

MeSH (Medical Subject Headings) helps you create networks based on information in PubMed abstracts and identify interactions associated MeSH terms using NLP as an alternate way of creating pathways based on terms and concepts instead of entities. Mass Profiler Professional obtains MeSH terms from the MeSH database (see http://www.nlm.nih.gov/mesh/meshhome.html).

1. Launch **MeSH Network Builder** in the Workflow Browser.

a Click **MeSH Network Builder** in the Workflow Browser.

The MeSH Network Builder operation has four (4) steps as shown in Figure 128. The new pathway list is placed in the Analysis folder within the Experiment Navigator. The organism used is the same as specified in the active project.

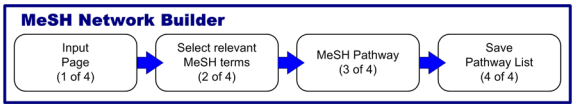


Figure 128 *Flow chart of the MeSH Network Builder operation.*

2. Input parameters in **MeSH Network Builder (Step 1 of 4)**.

a Type an **MeSH Term**.

Type a concept or actual MeSH term. The term does not have to be technical; a simple phrase or phenomenon of interest is sufficient, for example “memory.”

b Click **Next**.

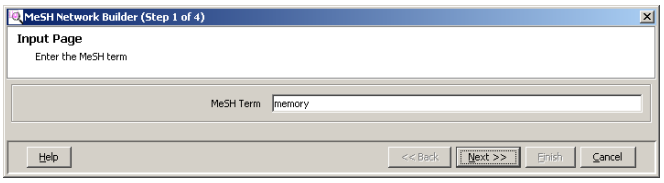


Figure 129 *Input Page (MeSH Network Builder (Step 1 of 4))*

3. Select terms in **MeSH Network Builder (Step 2 of 4)**.

a Mark the relevant MeSH headings the contain your input term(s).

b Select the filtering option for the relevant MeSH terms under **Select Type**.

Exact Relations includes only those interactions that contain the exact MeSH headings that were selected.

All Relevant Relations includes all interactions that contain either the exact MeSH heading or the child MeSH heading terms.

c Type in the value for **Min Frequency**.

Min Frequency is the minimum number of PubMed articles (PMIDs) associated with the MeSH term that an interaction should have. For example, if the Min Frequency is set to 5 the pathway includes only those interactions which have at least 5 PMIDs that contain the relevant MeSH term. The default setting value is 1.

d Click **Next**.

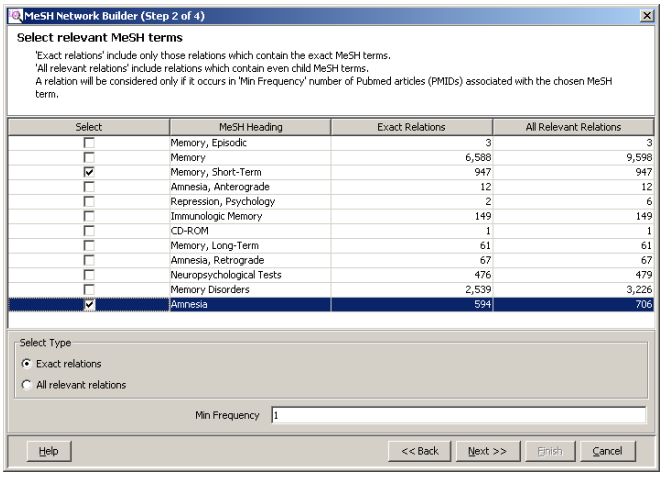

MeSH Network Builder (Step 2 of 4)
Select relevant MeSH terms
'Exact relations' include only those relations which contain the exact MeSH terms.
'All relevant relations' include relations which contain even child MeSH terms.
A relation will be considered only if it occurs in 'Min Frequency' number of Pubmed articles (PMIDs) associated with the chosen MeSH term.
Select
MeSH Heading
Exact Relations
All Relevant Relations
Memory, Episodic 3 3
Memory 6,588 9,598
Memory, Short-Term 947 947
Amnesia, Anterograde 12 12
Repression, Psychology 2 6
Immunologic Memory 149 149
CD-ROM 1 1
Memory, Long-Term 61 61
Amnesia, Retrograde 67 67
Neuropsychological Tests 476 479
Memory Disorders 2,539 3,226
Amnesia 594 706
Select Type
Exact relations
All relevant relations
Min Frequency
1
Help
<< Back
Next >>
Finish
Cancel

Figure 130 *Select relevant MeSH terms page (MeSH Network Builder (Step 2 of 4))*

4. Review the MeSH pathway in **MeSH Network Builder (Step 3 of 4)**.

MeSH Pathway displays the created pathway. The number of entities and the number of relations are displayed.

a Review the pathway.

b Edit the pathway. Details for using the pathway view is described in section "11.1.3 Creating and Editing Pathways" in the *Mass Profiler Professional User Manual*.

c Click **Next**.

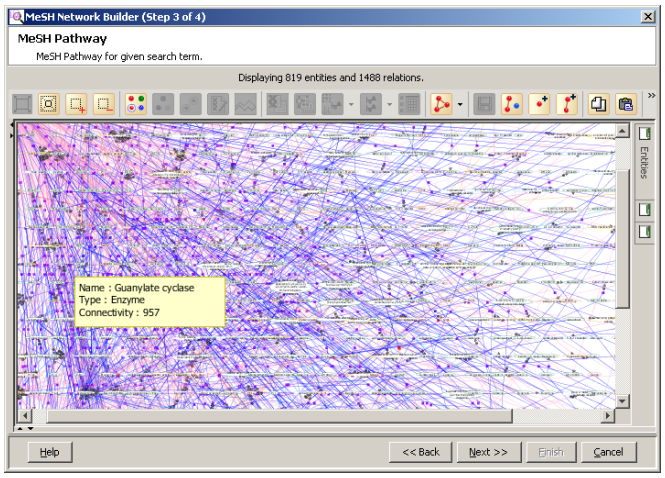


Figure 131 *MeSH Pathway page (MeSH Network Builder (Step 3 of 4))*

5. Save the pathway list in **MeSH Network Builder (Step 4 of 4)**.

a Review the pathway list.

b Type a descriptive **Name** that is stored with the saved pathway entity list.

c Edit the **Notes** that are stored with the saved pathway entity list.

d Double-click a row in the **Pathways** table to launch the Pathway Inspector to review the entities and relations contained in the new pathway.

e Click **Finish**.

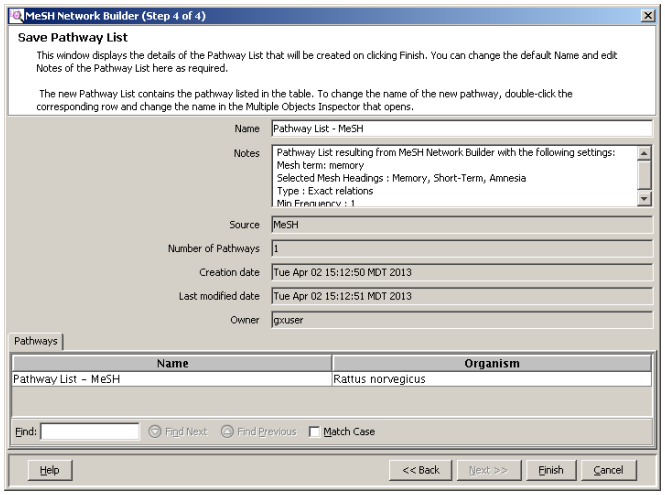

MeSH Network Builder (Step 4 of 4)
Save Pathway List
This window displays the details of the Pathway List that will be created on clicking Finish. You can change the default Name and edit Notes of the Pathway List here as required.
The new Pathway List contains the pathway listed in the table. To change the name of the new pathway, double-click the corresponding row and change the name in the Multiple Objects Inspector that opens.
Name
Pathway List - MeSH
Notes
Pathway List resulting from MeSH Network Builder with the following settings:
Mesh term: memory
Selected Mesh Headings : Memory, Short-Term, Amnesia
Type : Exact relations
Source
MeSH
Number of Pathways
1
Creation date
Tue Apr 02 15:12:50 MDT 2013
Last modified date
Tue Apr 02 15:12:51 MDT 2013
Owner
gxuser
Pathways
Name
Organism
Pathway List - MeSH
Rattus norvegicus
Find:
Find Next
Find Previous
Match Case
Help
<< Back
Next >>
Finish
Cancel

Figure 132 *Save Pathway List page (MeSH Network Builder (Step 4 of 4))*

Extract Relations via NLP

You can use Natural Language Processing (NLP) to create new pathways directly from PubMed abstracts and other documents stored on your PC or network (.pdf, .doc, and .html files). NLP first recognizes entities in sentences and then performs information extraction to identify relationships between these entities. The entities that NLP can recognize are restricted to those packaged in the generic Interaction Database and the Interaction Databases for the organism of the currently active experiment

1. Check that the **NLP limits** is greater than 1000.

a Click **Tools > Options** on the menu bar to launch configuration options.

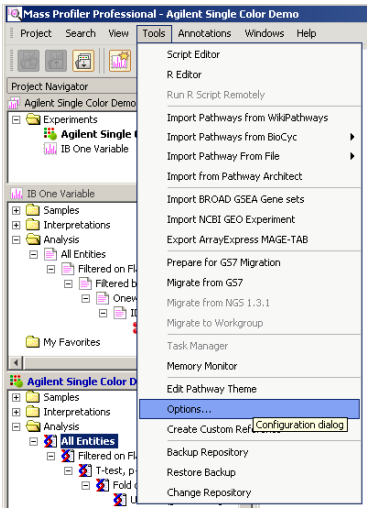


Figure 133 *Launching the configuration options from the menu bar*

b Click **Pathway** on the left-hand pane in the **Configuration Dialog** dialog box.

c Click **NLP limits** on the left-hand pane in the **Configuration Dialog** dialog box.

d Type a value lager than `1000` in the **Maximum no. Pubmed's to search** for the **NLP limits**.

Note: If the **NLP limit** value is smaller than 1000 you may receive an error "No abstracts found on PubMed" and "Cannot Process Input File" when you launch Extract Relations via NLP.

e Click **OK**.

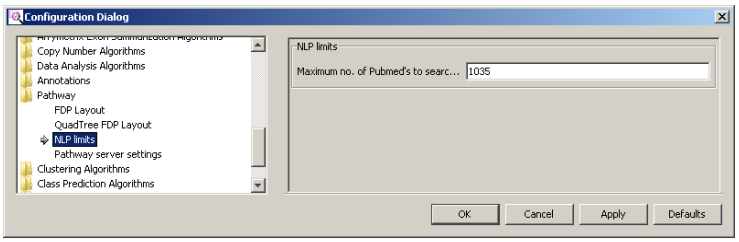


Figure 134 *NLP limits in the Configuration Dialog*

2. Launch **Extract Relations via NLP** in the Workflow Browser.

a Click **Extract Relations via NLP** in the Workflow Browser.

Extract Relations via NLP four (4) steps as shown in Figure 135. The new pathway list is placed in the Analysis folder within the Experiment Navigator.

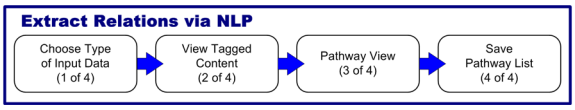

Extract Relations via NLP
Choose Type of Input Data (1 of 4)
View Tagged Content (2 of 4)
Pathway View (3 of 4)
Save Pathway List (4 of 4)

Figure 135 *Flow chart of the Extract Relations via NLP operation.*

3. Choose input parameters in **Extract Relations via NLP (Step 1 of 4)**.

a Select the **Input source**.

If the chosen **Input source** is "PubMed search" you can specify a search query that is submitted to PubMed. You can also choose to run NLP on your local files. Non-text format files such as .doc and .pdf files are converted into text using publicly available converters.

b Type in your **Search text**, `memory`, or click **Choose files** depending on your **Input source**.

c Click **Next**.

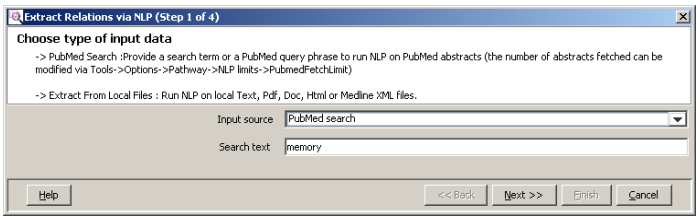

Extract Relations via NLP (Step 1 of 4)
Choose type of input data
-> PubMed Search :Provide a search term or a PubMed query phrase to run NLP on PubMed abstracts (the number of abstracts fetched can be modified via Tools->Options->Pathway->NLP limits->PubmedFetchLimit)
-> Extract From Local Files : Run NLP on local Text, Pdf, Doc, Html or Medline XML files.
Input source
PubMed search
Search text
memory
Help
<< Back
Next >>
Finish
Cancel

Figure 136 *Choose Type of Input Data page (Extract Relations via NLP (Step 1 of 4))*

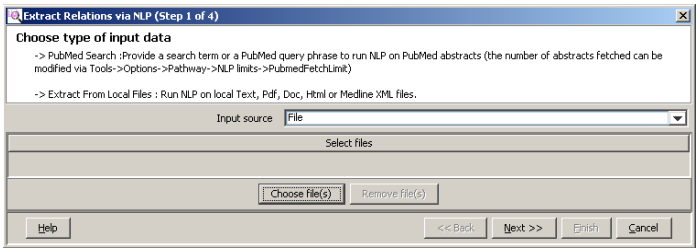

Extract Relations via NLP (Step 1 of 4)
Choose type of input data
-> PubMed Search :Provide a search term or a PubMed query phrase to run NLP on PubMed abstracts (the number of abstracts fetched can be modified via Tools->Options->Pathway->NLP limits->PubmedFetchLimit)
-> Extract From Local Files : Run NLP on local Text, Pdf, Doc, Html or Medline XML files.
Input source
File
Select files
Choose file(s)
Remove file(s)
Help
<< Back
Next >>
Finish
Cancel

Figure 137 *Choose Type of Input Data page for File Input source (Extract Relations via NLP (Step 1 of 4))*

4. Review tagged content in **Extract Relations via NLP (Step 2 of 4)**.

a Review the tagged content.

Target documents containing the search terms are identified. Tagging is performed using the entity dictionary. All molecular and biological processes and function entities present in the Mass Profiler Professional Interaction Databases are tagged. Matching entities are highlighted according to default color settings, with the corresponding legend displayed below the searched content. In the case of PubMed articles (or Medline XML files), the PubMed ID is shown in the left hand column. In all other cases, the name of the file is displayed.

b Click **Next**.

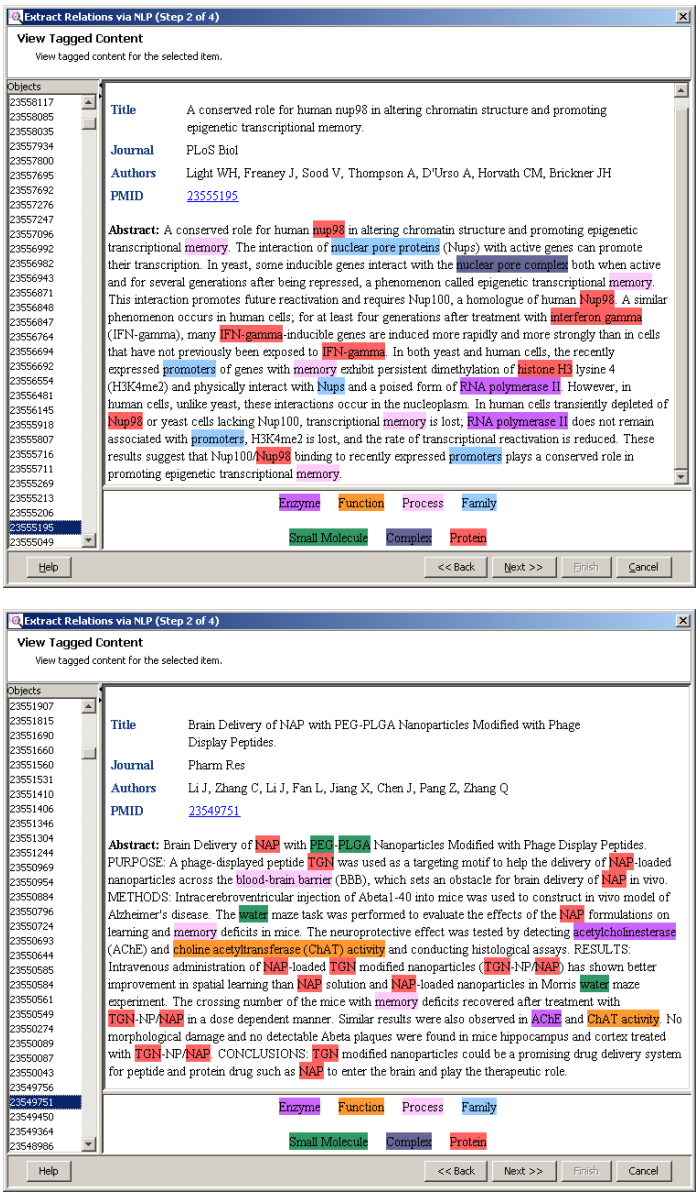

Extract Relations via NLP (Step 2 of 4)
View Tagged Content
View tagged content for the selected item.
Objects
Title
A conserved role for human nup98 in altering chromatin structure and promoting epigenetic transcriptional memory.
Journal
PLoS Biol
Authors
Light WH, Freaney J, Sood V, Thompson A, D'Urso A, Horvath CM, Brickner JH
PMID
23555195
Abstract: A conserved role for human nup98 in altering chromatin structure and promoting epigenetic transcriptional memory. The interaction of nuclear pore proteins (Nups) with active genes can promote their transcription. In yeast, some inducible genes interact with the nuclear pore complex both when active and for several generations after being repressed, a phenomenon called epigenetic transcriptional memory. This interaction promotes future reactivation and requires Nup100, a homologue of human Nup98. A similar phenomenon occurs in human cells; for at least four generations after treatment with interferon gamma (IFN-gamma), many IFN-gamma-inducible genes are induced more rapidly and more strongly than in cells that have not previously been exposed to IFN-gamma. In both yeast and human cells, the recently expressed promoters of genes with memory exhibit persistent dimethylation of histone H3 lysine 4 (H3K4me2) and physically interact with Nups and a poised form of RNA polymerase II. However, in human cells, unlike yeast, these interactions occur in the nucleoplasm. In human cells transiently depleted of Nup98 or yeast cells lacking Nup100, transcriptional memory is lost, RNA polymerase II does not remain associated with promoters, H3K4me2 is lost, and the rate of transcriptional reactivation is reduced. These results suggest that Nup100/Nup98 binding to recently expressed promoters plays a conserved role in promoting epigenetic transcriptional memory.
Enzyme
Function
Process
Family
Small Molecule
Complex
Protein
Help
<< Back
Next >>
Finish
Cancel
Extract Relations via NLP (Step 2 of 4)
View Tagged Content
View tagged content for the selected item.
Objects
Title
Brain Delivery of NAP with PEG-PLGA Nanoparticles Modified with Phage Display Peptides.
Journal
Pharm Res
Authors
Li J, Zhang C, Li J, Fan L, Jiang X, Chen J, Pang Z, Zhang Q
PMID
23549751
Abstract: Brain Delivery of NAP with PEG-PLGA Nanoparticles Modified with Phage Display Peptides. PURPOSE: A phage-displayed peptide TGN was used as a targeting motif to help the delivery of NAP-loaded nanoparticles across the blood-brain barrier (BBB), which sets an obstacle for brain delivery of NAP in vivo. METHODS: Intracerebroventricular injection of Abeta1-40 into mice was used to construct in vivo model of Alzheimer's disease. The water maze task was performed to evaluate the effects of the NAP formulations on learning and memory deficits in mice. The neuroprotective effect was tested by detecting acetylcholinesterase (AChE) and choline acetyltransferase (ChAT) activity and conducting histological assays. RESULTS: Intravenous administration of NAP-loaded TGN modified nanoparticles (TGN-NP/NAP) has shown better improvement in spatial learning than NAP solution and NAP-loaded nanoparticles in Morris water maze experiment. The crossing number of the mice with memory deficits recovered after treatment with TGN-NP/NAP in a dose dependent manner. Similar results were also observed in AChE and ChAT activity. No morphological damage and no detectable Abeta plaques were found in mice hippocampus and cortex treated with TGN-NP/NAP. CONCLUSIONS: TGN modified nanoparticles could be a promising drug delivery system for peptide and protein drug such as NAP to enter the brain and play the therapeutic role.
Enzyme
Function
Process
Family
Small Molecule
Complex
Protein
Help
<< Back
Next >>
Finish
Cancel

Figure 138 *Two target documents in the View Tagged Content page (Extract Relations via NLP (Step 2 of 4)) for memory as the search text*

5. Review the pathway in **Extract Relations via NLP (Step 3 of 4)**.

The Pathway View displays the created pathway. The number of relations are displayed above the pathway.

a Review the pathway.

b Edit the pathway. Details for using the pathway view is described in section "11.1.3 Creating and Editing Pathways" in the *Mass Profiler Professional User Manual*.

c Click **Next**.

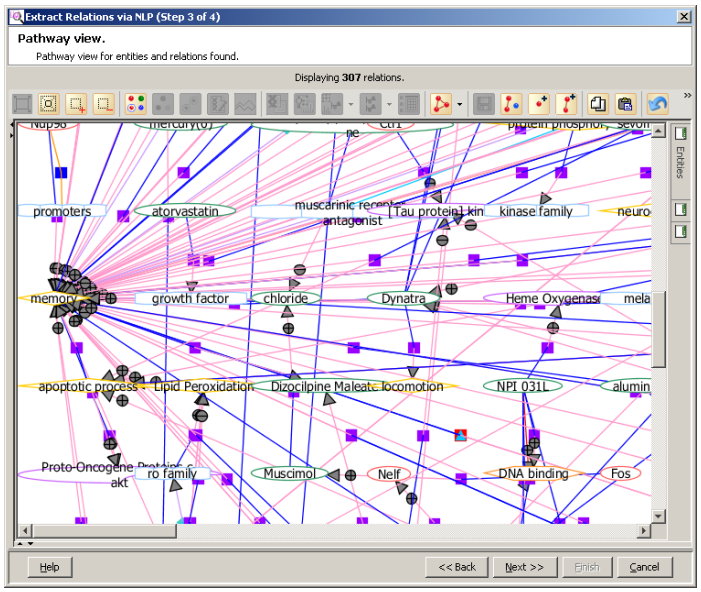


Figure 139 *Pathway View page (Extract Relations via NLP (Step 3 of 4)) for memory as the search text*

6. Save the pathway list in **Extract Relations via NLP (Step 4 of 4)**.

a Review the pathway list.

b Type a descriptive **Name** that is stored with the saved pathway entity list.

c Edit the **Notes** that are stored with the saved pathway entity list.

d Double-click a row in the **Pathways** table to launch the Pathway Inspector to review the entities and relations contained in the new pathway.

e Click **Finish**.

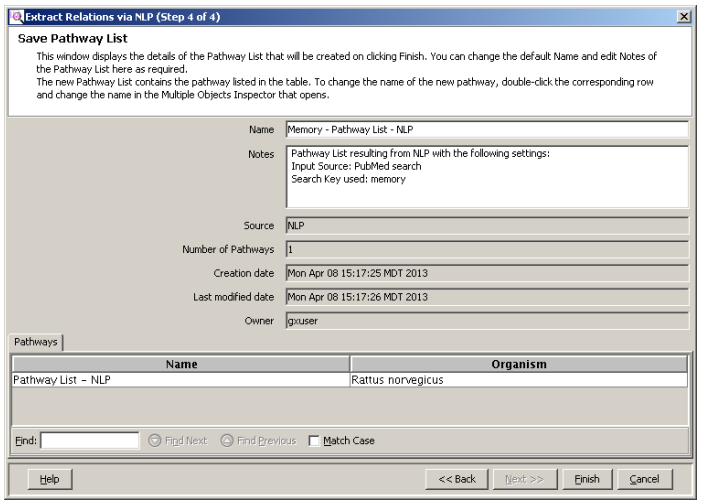

Extract Relations via NLP (Step 4 of 4)
Save Pathway List
This window displays the details of the Pathway List that will be created on clicking Finish. You can change the default Name and edit Notes of the Pathway List here as required.
The new Pathway List contains the pathway listed in the table. To change the name of the new pathway, double-click the corresponding row and change the name in the Multiple Objects Inspector that opens.
Name
Memory - Pathway List - NLP
Notes
Pathway List resulting from NLP with the following settings:
Input Source: PubMed search
Search Key used: memory
Source
NLP
Number of Pathways
1
Creation date
Mon Apr 08 15:17:25 MDT 2013
Last modified date
Mon Apr 08 15:17:26 MDT 2013
Owner
gxuser
Pathways
Name
Organism
Pathway List – NLP
Rattus norvegicus
Find:
Match Case
Help
<< Back
Next >>
Finish
Cancel

Figure 140 *Save Pathway List page (Extract Relations via NLP (Step 4 of 4))*

Create Pathway Organism

If you are working with an organism that is not currently available in Mass Profiler Professional you can create a new pathway organism.

a Click **Annotations > Create Pathway Organism**.

a Open http://www.ncbi.nlm.nih.gov/taxonomy in your Internet browser. Valid taxonomy IDs can be found in the Entrez Taxonomy database located at this site.

b Find the taxonomy for your organism. Information displayed on this page is entered into the **Create Pathway Organism** dialog box.

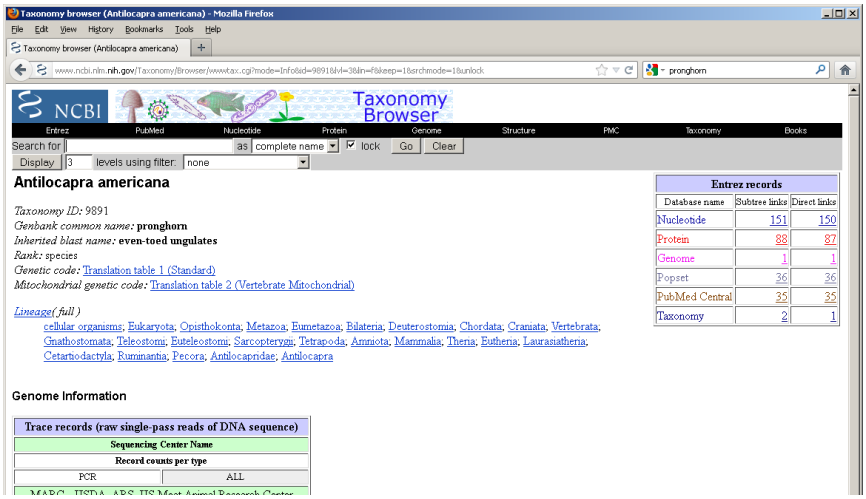


Figure 141 *The Entrez Taxonomy database for a North American mammal*

c Type the value for the **Taxonomy ID**.

d Type the exact **Scientific Name**, including capitalization and spaces.

e Type the **Common Name** using in your own style and spelling.

f Click **OK**. A progress box is displayed while the organism is added to Mass profiler Professional.

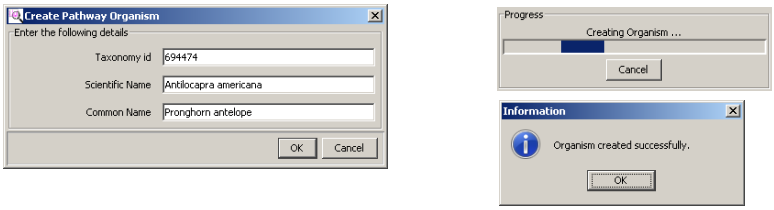


Figure 142 *Create Pathway Organism dialog box, organism creation progress, and Information dialog box indicating the successful creation of your organism*

g Click **OK** if you receive a notification that the organism is already supported in Mass Profiler Professional (see Figure 143 on page 107).

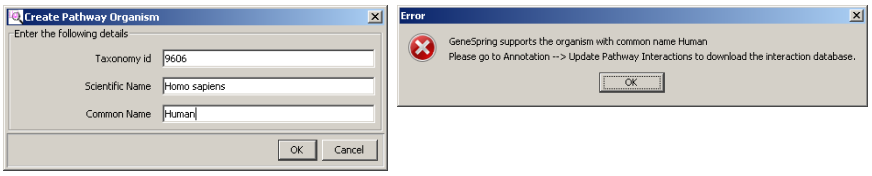


Figure 143 *Error indicating that the organism is already supported*

h Click **Annotations > Pathway Database Statistics** to confirm that your organism is now included in Mass Profiler Professional.

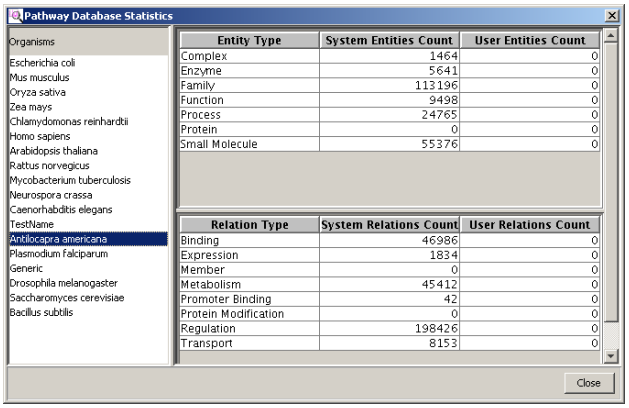
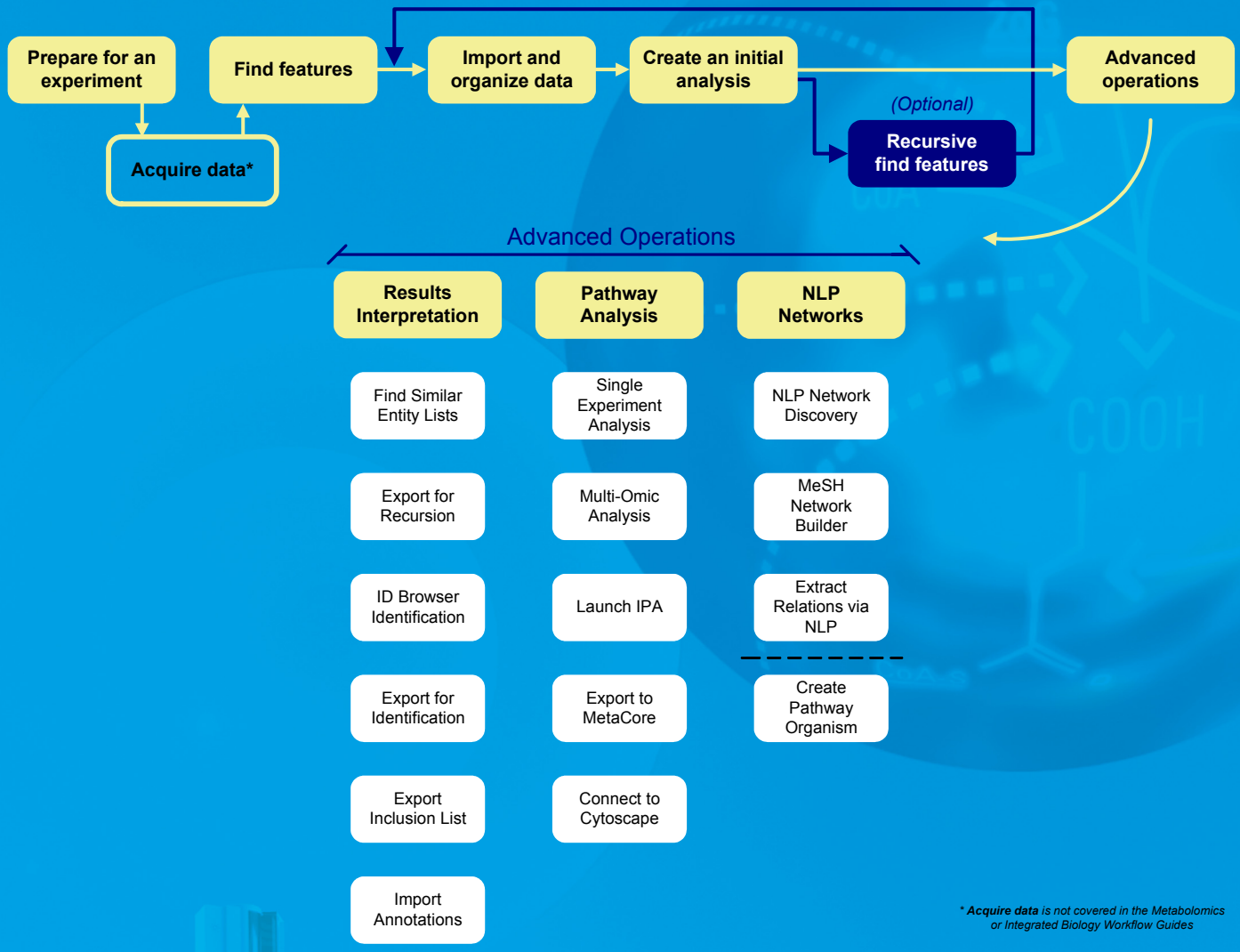


Figure 144 *Pathway Database Statistics information*

i Click **Close.** This completes creating a pathway organism.

Reference information

This chapter consists of definitions and references. The definitions section includes a list of terms and their definitions as used in this workflow. The references section includes citations to Agilent publications that help you use Agilent products and perform your metabolomics analyses.



** **Acquire data** is not covered in the Metabolomics or Integrated Biology Workflow Guides*

Definitions 110
References 120

Agilent Technologies

Definitions

This section contains a list of terms and their definitions as used in this workflow. Review of the terms and definitions presented in this section helps you understand the Agilent software wizards and the metabolomics workflow.

Alignment

Adjustment of the chromatographic retention time of eluting components to improve the correlation among data sets, based on the elution of specific component(s) that are (1) naturally present in each sample or (2) deliberately added to the sample through spiking the sample with a known compound or set of compounds that does not interfere with the sample.

AMDIS

Acronym for automated mass spectral deconvolution and identification system developed by NIST (http://www.amdis.net).

Amino acid

Biologically significant molecules that contain a core carbon positioned between a carboxyl and amine group in addition to an organic substituent. Dual carboxyl and amine functionalities facilitate the formation of peptides and proteins.

ANOVA

Abbreviation for analysis of variance which is a statistical method that simultaneously compares the means between two or more attributes or parameters of a data set. ANOVA is used to determine if a statistical difference exists between the means of two or more data sets and thereby prove or disprove the hypothesis. See also t-Test.

Attribute

Another term for an independent variable. Referred to as a parameter and is assigned a parameter name during the various steps of the metabolomic data analysis.

Attribute value

Another term for one of several values within an attribute for which exist correlating samples. Referred to as a condition or a parameter value and given an assigned value during the various steps of the metabolomic data analysis.

Baselining

A technique used to view and compare data that involves converting the original data values to values that are expressed as changes relative to a calculated statistical value derived from the data. The calculated statistical value is referred to as the baseline.

Bayesian

A term used to refer to statistical techniques named after the Reverend Thomas Bayes (ca. 1702 - 1761).

Bayesian inference

The use of statistical reasoning, instead of direct facts, to calculate the probability that a hypothesis may be true. Also known as Bayesian statistics.

Bioinformatics

The use of computers, statistics, and informational techniques to increase the understanding of biological processes.

Biomarker	An organic molecule whose presence and concentration in a biological sample indicates a normal or altered function of higher level biological activity.
Carbohydrate	An organic molecule consisting entirely of carbon, hydrogen, and oxygen that is important to living organisms.
CEF file	A binary file format called a compound exchange file (CEF) that is used to exchange data between Agilent software. In the metabolomics workflow CEF files are used to share molecular features between MassHunter Qualitative Analysis and Mass Profiler Professional.
Cell	The fundamental unit of an organism consisting of several sets of biochemical functions within an enclosing membrane. Animals and plants are made of one or more cells that combine to form tissues and perform living functions.
Census	Collection of a sample from every member of a population.
Cheminformatics	The use of computers and informational techniques (such as analysis, classification, manipulation, storage, and retrieval) to analyze and solve problems in the field of chemistry.
Chemometrics	A science employing mathematical and analytical processes to extract information from chemical data sets. The processes involve interactive applications of techniques employed in disciplines such as multivariate statistics, applied mathematics, and computer science to obtain meaningful information from complex data sets. Chemometrics is typically used to obtain meaningful information from data derived from chemistry, biochemistry and chemical engineering. Agilent Mass Profiler Professional is designed to employ chemometrics processes to GC/MS and LC/MS data sets to obtain useful information.
Child	A subset of information that is created by an algorithm from an original set of information. An entity list created using Mass Profiler Professional is a child. An original entity list is referred to as the parent of one or more child entity lists.
Co-elution	When compounds elute from a chromatographic column at nominally the same time making the assignment of the observed ions to each compound difficult.
Complex	Class of compounds consisting of more than one protein physically which physically bind each other and are biologically active and stable in their combined form.
Composite spectrum	A compound spectrum generated to represent the molecular feature that includes more than one ion, isotope, or adduct (not just M + H) and is used by Mass Profiler Professional for recursive analysis and ID Browser.

Compound	A metabolite that may be individually referred to as a compound, molecular feature, element, or entity during the various steps of the metabolomic data analysis.
Condition	Another term for one of several values within a parameter for which exist correlating samples. Condition may also be referred to as a parameter value during the various steps of the metabolomic data analysis. See also attribute value.
Data	Information in a form suitable for storing and processing by a computer that represent the qualitative or quantitative attributes of a subject. Examples include GC/MS and LC/MS data consisting fundamentally of time, ion m/z, and ion abundance from a chemical sample.
Data processing	Conversion of data into meaningful information. Computers are employed to enable rapid recording and handling of large amounts of data, i.e. Agilent MassHunter Workstation and Agilent Mass Profiler Professional.
Data reduction	See reduction.
Deconvolution	The technique of reconstructing individual mass and mass spectral data from co-eluting compounds.
Dependent variable	An element in a data set that can only be observed as a result of the influence from the variation of an independent variable. For example, a pharmaceutical compound structure and quantity may be controlled as two independent variables while the metabolite profile presents a host of small-molecule products that make up the dependent variables of a study.
Determinate	Having exact and definite limits on an analytical result that provide a conclusive degree of correlation of the subject to the specimen.
Element	A metabolite that may be individually referred to as a compound, molecular feature, element, or entity during the various steps of the metabolomic data analysis.
Endogenous	Pertaining to cause, development, or origination from within an organism.
Entity	A metabolite that may be individually referred to as a compound, molecular feature, element, or entity during the various steps of the metabolomic data analysis.
Entity List	The compounds that meet the requirements specified by each experiment performed on your data. Entity lists are viewed in the Experiment Navigator.
Enzyme	Proteins acting as biocatalysts in a metabolomic reaction. These entities are particularly important in depicting a biochemical network.

Experiment	Data acquired in an attempt to understand causality where tests or analyses are defined and performed on an organism to discover something that is not yet known, to demonstrate as proof of something that is known, or to find out whether something is effective.
Externality	A quality, attribute, or state that originates and/or is established independently from the specimen under evaluation.
Extraction	The process of retrieving a deliberate subset of data from a larger data set whereby the subset of the data preserves the meaningful information as opposed to the redundant and less meaningful information. Also known as data extraction.
Family	A group of proteins related by structure, function, or another biological parameter.
Feature	Independent, distinct characteristic of a phenomena and data under observation. Features are an important part of the identification of patterns - pattern recognition - within data whether processed by a human or by artificial intelligence, such as Agilent MassHunter Workstation and Agilent Mass Profiler Professional. In metabolomics analysis a feature is a metabolite and may be individually referred to as a compound, molecular feature, element, or entity during the various steps of the metabolomic data analysis.
Feature extraction	The reduction of data size and complexity through the removal of redundant and non-specific data by using the important variables (features) associated with the data. Careful feature extraction yields a smaller data set that is more easily processed without any compromise in the information quality. This is part of the principal component analysis process employed by Agilent Mass Profiler Professional.
Feature selection	The identification of important, or non-important, variables and the variable relationships in a data set using both analytical and a priori knowledge about the data. This is part of the principal component analysis process employed by Agilent Mass Profiler Professional.
Filter	The process of establishing criteria by which entities are removed (filtered) from further analysis during the metabolomics workflow.
Filter by flag	A flag is a term used to denote a quality of an entity within a sample. A flag indicates if the entity was detected in each sample as follows: Present means the entity was detected, Absent means the entity was not detected, and Marginal means the signal for the entity was saturated
Function	A classification of compounds based on their biological purpose or activity.
Hypothesis	A proposition made to explain certain facts and tentatively accepted to provide a basis for further investigation. A proposed explanation for observable phenomena

	may or may not be supported by the analytical data. Statistical data analysis is performed to quantify the probability that the hypothesis is true. Also known as the scientific hypothesis.
Hypothetical	A statement based on, involving, or having the nature of a hypothesis for the purposes of serving as an example and not necessarily based on an actuality.
ID Browser	Agilent software that automatically annotates the entity list with the compound names and adds them to any of the various visualization and pathway analysis tools.
Identified compound	Chromatographic components that have an assigned, exact identity, such as compound name and molecular formula, based on prior assessment or comparison with a database. See also Unidentified Compound.
Independent variable	An essential element, constituent, attribute, or quality in a data set that is deliberately controlled in an experiment. For example, a pharmaceutical compound structure and quantity may be controlled as two independent variables while the metabolite profile presents a host of independent small molecule products that make up the dependent variables of a study. An independent variable may be referred to as a parameter and is assigned a parameter name during the various steps of the metabolomic data analysis.
Inorganic compound	Non carbon and non biological origin compounds such as minerals and salts.
Interpretation	Expression of your data in entity lists after grouping your samples, applying filters, and performing statistical correlation methods. When you open an experiment, the "All Samples" interpretation is active. You can click on another interpretation to activate it.
Lipidomics	Identification and quantification of cellular lipids from an organism in a specified biological situation. The study of lipids is a subset of metabolomics.
Mass variation	Using the mass to charge (m/z) resolution to improve compound identification. Compounds with nearly identical and identical chromatographic behavior are deconvoluted by adjusting the m/z range for extracting ion chromatograms.
Mean	The numerical result of dividing the sum of the data values by the number of individual data observations.
Metabolism	The chemical reactions and physical processes whereby living organisms convert ingested compounds into other compounds, structures, energy and waste.
Metabolite	Small organic molecules that are intermediate compounds and products produced as part of metabolism. Metabolites are important modulators, substrates, byproducts, and building blocks of many different biological processes. In order to distin-

guish metabolites from lager biological molecules, known as macromolecules such as proteins, DNA and others, metabolites are typically under 1000 Da. A metabolite may be individually referred to as a compound, molecular feature, element, or entity during the various steps of the metabolomic data analysis.

Metabolome

The complete set of small-molecule metabolites that may be found within a biological sample. Small molecules are typically in the range of 50 to 600 Da.

Metabolomics

The process of identification and quantification of all metabolites of an organism in a specified biological situation. The study of the metabolites of an organism presents a chemical "fingerprint" of the organism under the specific situation. See metabonomics for the study of the change in the metabolites in response to externalities.

Metabonomics

The metabolic response to externalities such as drugs, environmental factors, and disease. The study of metabonomics by the medical community may lead to more efficient drug discovery and to individualized patient treatment. Meaningful information learned from the metabolite response can be used for clinical diagnostics or for understanding the onset and progression of human diseases. See metabolomic for the identification and quantitation of metabolites.

NLP

Natural Language Processing (NLP) algorithm that extracts information from published literature.

Normalization

A technique used to adjust the ion intensity of mass spectral data from an absolute value based on the signal measured at the detector to a relative intensity of 0 to 100 percent based on the signal of either (1) the ion of the greatest intensity or (2) a specific ion in the mass spectrum.

Null hypothesis

The default position taken by the hypothesis that no effect or correlation of the independent variables exists with respect to the measurements taken from the samples.

Observation

Data acquired in an attempt to understand causality where no ability exists to (1) control how subjects are sampled and/or (2) control the exposure each sample group receives.

One-hit wonder

An entity that appears in only one sample, is absent from the replicate samples, and does not provide any utility for statistical analysis. Entities that are one-hit wonders may be filtered using Filter by Flags.

Organic compound

Carbon-based compounds, often with biological origin.

Organism

A group of biochemical systems that function together as a whole thereby creating an individual living entity such as an animal, plant, or microorganism. Individual living entities may be multicellular or unicellular. See also specimen.

p-value

The probability of obtaining a statistical result that is comparable to or greater in magnitude than the result that was actually observed, assuming that the null hypothesis is true. The null hypothesis is stated that no correlation exists between the independent variables and the measurements taken from the samples. Rejection of the null hypothesis is typically made when the p-value is less than 0.05 or 0.01. A p-value of 0.05 or 0.01 may be restated as a 5% or 1% chance of rejecting the null hypothesis when it is true. When the null hypothesis is rejected, the result is said to be statistically significant meaning that a correlation exists between the independent variables and the measurements as specified in the hypothesis.

Parameter

Another term for an independent variable. Referred to as a parameter or parameter name and is assigned a parameter name during the various steps of the metabolomic data analysis. See also condition and attribute.

Parameter value

Another term for one of several values within a parameter for which exist correlating samples. Parameter value may also be referred to as a condition during the various steps of the metabolomic data analysis. See also attribute value.

Parent

The original set of information that is processed by an algorithm to create one or more subsets of information. A subset entity list is referred to as the child of a parent entity list.

Peptide

Linear chain of amino acids that is shorter than a protein. The length of a peptide is sufficiently short that it is easily made synthetically from the constituent amino acids.

Peptide bond

The covalent bond formed by the reaction of a carboxyl group with an amine group between two molecules, e.g. between amino acids.

Permutation

Any of the total number of subsets that may be formed by the combination of individual parameters among the independent variables. For example the number of permutations of A and B in variable Φ in combination with X, Y, and Z in variable θ equals six (6 = 2 x 3) and may be represented as AX, AY, AZ, BX, BY, and BZ. Note that the combinations of parameters within a variable are not relevant such as AB, XY, XZ, and YZ.

Polarity

The condition of an effect as being positive or negative, additive or subtractive, with respect to some point of reference, such as with respect to the concentration of a metabolite.

Polymer

A molecule formed by the covalent bonding of a repeating molecular group to form a larger molecule.

Pooled sample

When the amount of available biological material is very small samples may be combined into a single sample (pooled) and then split into different aliquots for multiple analyses. By pooling the sample, sufficient material exists to obtain replicate analy-

ses of each sample where formerly there was insufficient material to obtain replicate analytical results. The trade-off loss of information about the biological variation that was formerly present in each unique sample is offset by a gain in statistical significance of the results.

Principal component

Transformed data into axes, or principal components, so that the patterns between the axes most closely describe the relationships between the data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The principal components often may be viewed, and interpreted, most readily in graphical axes with additional dimensions represented by color and/or shape representing the key elements (independent variables) of the hypothesis. This is part of the principal component analysis process employed by Agilent Mass Profiler Professional.

Principal component analysis

The mathematical process by which data containing a number of potentially correlated variables is transformed into a data set in relation to a smaller number of variables called principal components which account for the most variability in the data. The result of the data transformation leads to the identification of the best explanation of the variance in the data, e.g. identification of the meaningful information. Also known as PCA.

Protein

Linear chain of amino acids whose amino acid order and three-dimensional structure are essential to living organisms. Also know as a polypeptide.

Proteomics

The study of the structure and function of proteins occurring in living organisms. Proteins are assemblies of amino acids (polypeptides) based on information encoded in the genes of an organism and are the main components of the physiological metabolic pathways of the organism.

Quality

A feature, attribute, and/or characteristic element whose presence, absence, or inability to be properly ascertained due to instrumental factors, is factored into whether a sample is or is not representative of the larger specimen.

Recursive

Reapplying the same algorithm to a subset of a previous result in order to generate an improved result.

Recursive finding

A three-step process in the metabolomics workflow that improves the accuracy of finding statistically significant features in sample data files. Step 1: Find untargeted compounds by molecular feature in MassHunter Qualitative Analysis. Step 2: Filter the molecular features in Mass Profiler Professional. Step 3: Find targeted compounds by formula in MassHunter Qualitative Analysis. Importing the most significant features identified using Mass Profiler Professional back into MassHunter Qualitative Analysis as targeted features improves the accuracy in finding these features from the original sample data files.

Reduction

The process whereby the number of variables in a data set is decreased to improve computation time and information quality. For example, an extracted ion chromatogram obtained from GC/MS and LC/MS data files. Reduction provides smaller, viewable and interpretable data sets by employing feature selection and feature extraction. Also know as dimension reduction and data reduction. This is part of the principal component analysis process employed by Agilent Mass Profiler Professional.

Regression analysis

Mathematical techniques for analyzing data to identify the relationship between dependent and independent variables present in the data. Information is gained from the estimation, regression, or the sign and proportionality of the effects of the independent variables on the dependent variables. This is part of the principal component analysis process employed by Agilent Mass Profiler Professional. Also known as regression.

Replicate

Collecting multiple identical samples from a population so that when the samples are evaluated a value is obtained that more closely approximates the true value.

Sample

A part, piece, or item that is taken from a specimen and understood as being representative of the larger specimen (e.g., blood sample, cell culture, body fluid, aliquot) or population. An analysis may be derived from samples taken at a particular geographical location, taken at a specific period of time during an experiment, or taken before or after a specific treatment. A small number of specimens used to represent a whole class or group.

Sample class prediction

A workflow used to build a model and classify samples from mass spectrometry data. Class prediction is a supervised learning method and involves three steps: validation, training, and prediction. The algorithm learns from samples (training set) with known functional class and builds a prediction model to classify new samples (test set) of unknown class.

Specimen

An individual organism, e.g., a person, animal, plant, or other organism, of a class or group that is used as a representative of a whole class or group.

Spike

The specific and quantitative addition of one or more compounds to a sample.

Standard

A chemical or mixture of chemicals selected for use as a basis of comparing the quality of analytical results or for use to measure and compensate the precise offset or drift incurred over a set of analyses.

Standard deviation

A measure of variability among a set of data that is equal to the square root of the arithmetic average of the squares of the deviations from the mean. A low standard deviation value indicates that the individual data tend to be very close to the mean, whereas a high standard deviation indicates that the data is spread out over a larger range of values from the mean.

State

A set of circumstances or attributes characterizing a biological organism at a given time. A few sample attributes may include temperature, time, pH, nutrition, geography, stress, disease, and controlled exposure.

Statistics

The mathematical process employed in manipulating numerical data from scientific experiments to derive meaningful information. This is part of the principal component analysis, t-test, and ANOVA processes employed by Agilent Mass Profiler Professional.

Subject

A chemical or biological sample taken from a specimen, or a whole specimen, that undergoes a treatment, experiment, or an analysis for the purposes of further understanding.

Survey

Collection of samples from less than the entire population in order to estimate the population attributes.

t-Test

A statistical test to determine whether the mean of the data differs significantly from that expected if the samples followed a normal distribution in the population. The test may also be used to assess statistical significance between the means of two normally distributed data sets. See also ANOVA.

Unidentified compound

Chromatographic components that are only uniquely denoted by their mass and retention times and which have not been assigned an exact identity, such as compound name and molecular formula. Unidentified compounds are typically produced by feature finding and deconvolution algorithms. See also Identified Compound.

Variable

An element in a data set that assumes changing values, e.g. values that are not constant over the entire data set. The two types of variables are independent and dependent.

Volume

The area of the extracted compound chromatogram (ECC). The ECC is formed from the sum of the individual ion abundances within the compound spectrum at each retention time in the specified time window. The compound volume generated by MFE is used by Mass Profiler Professional to make quantitative comparisons.

Wizard

A sequence of dialog boxes presented by Mass Profiler Professional that guides you through well-defined steps to enter information, organize data, and perform analyses.

References

This section consists of citations to Agilent manuals, primers, application notes, presentations, product brochures, technical overviews, training videos, and software that help you use Agilent products and perform your metabolomics analyses.

Manuals

- Agilent G3835AA MassHunter Mass Profiler Professional - Quick Start Guide
  (Agilent publication, G3835-90009, Revision A, November 2012)
- Agilent G3835AA MassHunter Mass Profiler Professional - Familiarization Guide
  (Agilent publication, G3835-90010, Revision A, November 2012)
- Agilent G3835AA MassHunter Mass Profiler Professional - Application Guide
  (Agilent publication, G3835-90011, Revision A, November 2012)
- Agilent Metabolomics Workflow - Discovery Workflow Guide
  (Agilent publication 5990-7067EN, Revision B, October 2012)
- Agilent Metabolomics Workflow - Discovery Workflow Overview
  (Agilent publication 5990-7069EN, Revision B, October 2012)
- Agilent Mass Profiler Professional -
  (Agilent publication, January 2012)
- Agilent MassHunter Workstation Software Qualitative Analysis - Familiarization Guide
  (Agilent publication G3336-90018, Revision A, September 2011)
- Agilent MassHunter Workstation Software Quantitative Analysis - Familiarization Guide
  (Agilent publication G3335-90108, First Edition, June 2011)

Primers

- Proteomics: Biomarker Discovery and Validation
  (Agilent publication 5990-5357EN, February 11, 2010)
- Metabolomics: Approaches Using Mass Spectrometry
  (Agilent publication 5990-4314EN, October 27, 2009)

Application Notes

- Multi-omic Analysis with Agilent's GeneSpring 11.5 Analysis Platform
  (Agilent publication 5990-7505EN, March 25, 2011)
- An LC/MS Metabolomics Discovery Workflow for Malaria-Infected Red Blood Cells Using Mass Profiler Professional Software and LC-Triple Quadrupole MRM Confirmation
  (Agilent publication 5990-6790EN, November 19, 2010)
- Profiling Approach for Biomarker Discovery using an Agilent HPLC-Chip Coupled with an Accurate-Mass Q-TOF LC/MS
  (Agilent publication 5990-4404EN, October 20, 2009)
- Metabolite Identification in Blood Plasma Using GC/MS and the Agilent Fiehn GC/MS Metabolomics RTL Library
  (Agilent publication 5990-3638EN, April 1, 2009)
- Metabolomic Profiling of Bacterial Leaf Blight in Rice
  (Agilent publication 5989-6234EN, February 14, 2007)

Presentations

- Advances in Instrumentation and Software for Metabolomics Research
  (Agilent publication n/a, September 18, 2012)
- Multi-omics Analysis Software for Targeted Identification of Key Biological Pathways
  (Agilent publication n/a, May 3, 2012)
- Metabolomics LCMS Approach to: Identifying Red Wines according to their variety and Investigating Malaria infected red blood cells
  (Agilent publication n/a, November 3, 2010)
- Small Molecule Metabolomics
  (Agilent publication n/a, November 3, 2010)
- Presentation: Metabolome Analysis from Sample Prep through Data Analysis
  (Agilent publication n/a, November 3, 2010)

Product Brochures

- Emerging Insights: Agilent Solutions for Metabolomics
  (Agilent publication 5990-6048EN, April 30, 2012)
- Agilent Mass Profiler Professional Software - Discover the Difference in your Data
  (Agilent publication 5990-4164EN, April 27, 2012)
- Pathways to Insight - Integrated Biology at Agilent
  (Agilent publication 5991-0222EN, March 30, 2012)
- Confidently Better Bioinformatics Solutions
  (Agilent publication 5990-9905EN, February 2, 2012)
- Integrated Biology from Agilent: The Future is Emerging
  (Agilent publication 5990-6047EN, September 1, 2010)
- Agilent Fiehn GC/MS Metabolomics RTL Library
  (Agilent publication 5989-8310EN, December 5, 2008)
- Agilent METLIN Personal Metabolite Database
  (Agilent publication 5989-7712EN, December 31, 2007)
- Agilent Metabolomics Laboratory: The breadth of tools you need for successful metabolomics research
  (Agilent publication 5989-5472EN, January 31, 2007)

BioCyc Pathway/Genome Databases

Includes BioCyc Pathway/Genome databases from the Bioinformatics Research Group at SRI International®, used under license.

http://www.biocyc.org/

Citation based on use of BioCyc

Users who publish research results in scientific journals based on use of data from the EcoCyc Pathway/Genome database should cite:

Keseler et al, Nucleic Acids Research 39:D583-90 2011.

Users who publish research results in scientific journals based on use of data from most other BioCyc Pathway/Genome databases should cite:

Caspi et al, Nucleic Acids Research 40:D742-53 2012.

In some cases, BioCyc Pathway/Genome databases are described by other specific publications that can be found by selecting the database and then going to the Summary Statistics pages under the Tools menu. The resulting page sometimes contains a citation for that database.

www.agilent.com

© Agilent Technologies, Inc. 2013
Revision A, June 2013

5991-1909EN

Agilent Technologies